



HAL
open science

Are high–frequency traders informed?

Panagiotis Anagnostidis, Patrice Fontaine, Christos Varsakelis

► **To cite this version:**

Panagiotis Anagnostidis, Patrice Fontaine, Christos Varsakelis. Are high–frequency traders informed?.
Economic Modelling, 2020, 93, pp.365-383. 10.1016/j.econmod.2020.08.013 . hal-03062831

HAL Id: hal-03062831

<https://hal.science/hal-03062831>

Submitted on 14 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Are High-Frequency Traders Informed?

Panagiotis Anagnostidis

*Institut Europlace de Finance, Palais Brongniart 28 place de la Bourse, 75002 Paris, France
European Financial Data Institute, CNRS UPS 3390 – 150 rue de la chimie – CS 40 700 – 38058 Grenoble Cedex 9,
France*

Patrice Fontaine*

*European Financial Data Institute, CNRS UPS 3390 – 150 rue de la chimie – CS 40 700 – 38058 Grenoble Cedex 9,
France*

Léonard de Vinci Pôle Universitaire, Research Center, 92916 Paris La Défense, France

Christos Varsakelis

Université catholique de Louvain, iMMC, 1348 Louvain-la-Neuve, Belgium

Abstract

Are high-frequency traders (HFTs) informed? To address this question, we examine HFTs' activity in the call auction environment, where speed-related trading is limited and signal processing capacity becomes more relevant. To model the call market, we consider the Kyle (1989) rational expectations framework for strategic trading. The test we propose for detecting informed HFTs in this market assesses potential deviations of the informativeness of HFTs' aggregate (net) demand, from the informativeness of the aggregate demand submitted by the rest of the traders. Data from the Euronext Paris preopening phase indicate that informed HFTs are present in the market just before the opening. Our results provide useful guidance for the assessment of the influence of HFTs' quotes on price quality, an important issue for market regulators and policy makers.

Keywords: High-Frequency Trading, Call Auction, Rational Expectations, Price Efficiency

JEL Classification: G1, G14

Patrice Fontaine and Panagiotis Anagnostidis acknowledge support from French State (Agence Nationale de la Recherche), Investissements d'avenir ANR-11-EQPX-006 and in the framework of the Trans-Atlantis Platform ANR-16-DATA-002-04. The authors would like to thank the two anonymous Referees and the Editor for providing constructive comments that have improved the quality of the manuscript. We are also grateful to Sonia Jimenez-Garcés, Terrence Hendershott, Carole Metais, Duc Khuong Nguyen, and Mark Seasholes, for useful comments on previous versions of the paper, as well as the seminar participants at the Finance and Growth Workshop (ANR FIRE) (2017), AFFI (2017, 2018), FMA (2018) and Dauphine University (2018). Lastly, we are indebted to Clément Michoudet for his valuable research assistance.

*Corresponding author

Email addresses: panagiotis.anagnostidis@eurofidai.org (Panagiotis Anagnostidis),
patrice.fontaine@eurofidai.org (Patrice Fontaine), christos.varsakelis@uclouvain.be (Christos Varsakelis)

1 Introduction

A well-documented property of High-Frequency Traders (HFTs) is their ability to implement trading strategies at ultra-high speeds. Due to their technological advantage, HFTs are able to react first to price changes (typically within milliseconds or microseconds), raising their profits by trading at the expense of slower (e.g., human) market participants (Jarrow and Protter, 2012; Chlistalla, 2012; Cespa and Foucault, 2013; Biais and Foucault, 2014; Biais et al., 2015; Foucault et al., 2016; Hirschey, 2019). This property of HFTs has led to the widely-popular argument that, in modern electronic markets, being faster (slower) than the other traders is equivalent to being informed (uninformed) (O’Hara, 2015). The question, however, remains as to whether HFTs are informed about the values of securities. Brogaard et al. (2019, p. 1626), for example, remark that: “*Given that HFTs rely only on public information in their trading algorithms, one can argue that all of their contribution to price discovery is due to public information. However, if HFTs’ algorithms better interpret public signals, then it is more difficult to characterize HFTs as incorporating purely public information*”. In this paper, we address the question of whether HFTs hold accurate information about a stock’s value, by examining their quoting activity in the opening call of the Paris stock market.¹

In the call auction environment, orders are aggregated for an extended period of time during which there is no actual trading. At the end of this order accumulation period, clearing takes place at a single auction price. Because bidding is only theoretical until the clearing time, order processing speed is irrelevant (fast arbitrage and directional trading strategies, as well as flash trading strategies, cannot be put into practice), while information processing capacity becomes most important. Most major economic news, including earnings, is released before the market opens (Gao et al., 2018). We demonstrate that HFTs’ quotes during the last minutes of the preopening phase and at the opening trade are highly informative about stocks’ values. Our results indicate that HFTs efficiently aggregate and process information released overnight and in the early morning, contributing to price discovery. While we find that slow traders’ (non-HFTs’) quotes also contain information, particularly

¹Fox et al. (2018, p. 3) define a trader informed about a stock’s value as follows: “*A person generates fundamental value information by gathering various bits of information that are publicly available or otherwise observable features of the world and analysing what has been gathered in a sophisticated way that enables a superior assessment of a stock’s cash flows than that implied by the current market price*”. To avoid confusion with insider trading (e.g., private knowledge of a future corporate event, such as a merger, or a large planned sale or purchase of shares), we treat informed traders, on average, as value traders, in line with the above definition. Note, however, that we cannot exclude the possibility that insider trading may occasionally occur during the opening auction.

prior to the opening call, these non-high-frequency trading quotes are significantly noisier than those of HFTs, suggesting that HFTs are superior to non-HFTs in processing public information. This ability thus introduces information asymmetry into the call market.

To answer our research question, are HFTs informed about assets' values, we derive a test for detecting informed HFTs in the call market. Our starting point is the one-shot noisy rational expectations equilibrium model of Kyle (1989), which describes a transparent automated market where a finite number of strategic informed and uninformed traders coexist with liquidity/noise traders. This framework is suitable for modeling call auction markets, where traders' participation is limited (compared to the continuous trading process) and strategic bidding may affect the market price.² For this paradigm, comparative static analysis predicts that price efficiency is positively correlated with the number of informed and uninformed traders and is inversely correlated with the number of noise traders.

In theory, the distinction between informed and uninformed traders is clear. In practice, although it is fairly straightforward to empirically evaluate the information content of call auction prices (Biais et al., 1999; Barclay and Hendershott, 2003), inference on trader-type quote informativeness is a laborious task. Informative price shocks along with uninformative price variations, occurring during the order accumulation period, make it hard to discern whether certain groups of traders do or do not hold accurate information about the values of securities. To identify informed HFTs, we propose to investigate the information content of trader-type crossing supply/demand schedules. Our approach is based on the simple intuition that if HFTs do hold information about the value of a security, then this information should be reflected in their aggregate net demand.

We exploit the Kyle (1989) framework to separate the three types of market participants (informed, uninformed and noise traders) into subgroups of HFTs and non-HFTs. Subsequently, alongside the equilibrium price p , we analytically calculate two theoretical prices, p_{HFT} and $p_{non-HFT}$, that are defined uniquely at the point where the aggregate net demand submitted by HFTs and non-HFTs, respectively, is zero; we keep all other market parameters constant at their equilibrium value (see, for example, Figure 1). Then, by calculating deviations in the information content of

²Other one-shot rational expectations frameworks are those of Hellwig (1980), Admati (1985), and Madhavan and Panchapagesan (2000). Similar to the Kyle (1989) framework, these models are extensions of the seminal work on rational expectations by Grossman and Stiglitz (1976) and Grossman and Stiglitz (1980). See, also, the discussion in Kalay and Wohl (2009) for the appropriateness of these frameworks for modeling the single call auction mechanism.

p_{HFT} from the information content of $p_{non-HFT}$ and p , we are able to derive a test for detecting the presence of informed HFTs in the call market.

The data set we employ to conduct our empirical test includes the Euronext Paris preopening order placement history and opening call auction trading history for the CAC 40 (Cotation Assistée en Continu) stocks in year 2013. This database contains information on whether an order or a trade (buy or sell side) is from an HFT, based on the ratio of the individual order lifetime over the average order lifetime for canceled orders, for each market participant’s unique identifier.³ In this way we are able to directly disentangle HFTs from non-HFTs in the preopening phase.

We reconstruct the order book and calculate three types of empirical preopening indicative and opening prices, $p_{FULLBOOK}$, p_{HFT} and $p_{non-HFT}$, which minimize the aggregate net demand submitted by all traders, by HFTs and by non-HFTs, respectively. We empirically examine the information content of the reconstructed indicative and auction prices using the unbiasedness regression technique proposed by [Biais et al. \(1999\)](#); this technique provides an estimate of the uncertainty of prices about the value of the stock (i.e., the inverse of price efficiency).⁴ We then use the estimated levels of informativeness of $p_{FULLBOOK}$, p_{HFT} , and $p_{non-HFT}$ to test for the existence of informed HFTs.

We find that the early preopening orders are, on average, rather noisy, including those submitted by HFTs. Minutes before the opening, however, informed HFTs are present in the call market, and the ratio of noise to informed trading within the subgroup of HFTs becomes substantially lower, compared to the subgroup of non-HFTs. As the [Kyle \(1989\)](#) framework predicts, the presence of informed HFTs enhances the informativeness of the market price, which becomes fully efficient at the opening. Post-opening tests indicate that the prices fail to reverse in the continuous session (up to one hour after the opening), confirming that the opening price was indeed informative.

Our paper contributes to the literature on the effect of HFTs on price quality by investigating a trading system in which investors are allowed to gather and process information for an extended intraday period (1 hour and 45 minutes) until the clearing time. To the best of our knowledge, only a few other studies have examined thus far the characteristics of HFTs’ quotes in call auctions (e.g.,

³The high-frequency trading classification is provided by the Autorité des Marchés Financiers (AMF) (see, [Megarbane et al. \(2017\)](#)).

⁴The unbiasedness regression methodology is routinely employed in the call auction literature for the estimation of price informativeness ([Madhavan and Panchapagesan, 2000](#); [Ciccotello and Hatheway, 2000](#); [Barclay and Hendershott, 2003](#); [Comerton-Forde and Rydge, 2006](#); [Moshirian et al., 2012](#); [Anagnostidis et al., 2015](#); [Baruch et al., 2017](#); [Boussetta et al., 2020](#)).

Bellia et al., 2017; Boussetta et al., 2020), and none has focused on the detection of informed HFTs.

The paper is organized as follows: Section 2 reviews the related literature. Section 3 develops our test for detecting informed HFTs in the call market. Section 4 presents our empirical results from the Paris opening call. Finally, Section 5 concludes the paper.

2 Related literature

2.1 Price informativeness and high-frequency trading

Trading on information not yet reflected in a stock's price is important for capital market efficiency, as informative shocks accelerate the convergence of prices to their true values, in equilibrium. Our paper is motivated by findings in the literature that suggest that HFTs' activity can have both positive and negative effects on price informativeness. Foucault et al. (2016) develop a model where fast speculators take advantage of their speed by trading in the direction of anticipated short-term price changes (high-frequency news), therefore improving price discovery. At the empirical level, studying both NASDAQ and New York Stock Exchange (NYSE) listed stocks, Brogaard et al. (2014) find that HFTs improve price discovery by trading in the direction of short-term permanent price changes (interpreted as information) and in the opposite direction of transitory pricing errors. It can be argued, however, that changes in assets' values within such short time intervals may be less meaningful. Jarrow and Protter (2012), for example, show that HFTs may potentially harm price efficiency by implementing momentum trading strategies based on short-lived common signals. More specifically, commonality and the speed of information processing may cause temporary misspricings (i.e., deviations from the efficient price) which HFTs exploit at the expense of slow traders.

Our analysis contributes to the literature by examining the call auction mechanism where trading is not allowed for an extended intraday period. Such an exercise allows us to explore the information content of HFTs' quotes within a controlled environment where signal processing capacity is of critical importance, whereas speed-related trading is less relevant.⁵

⁵Other empirical findings in the literature that are indirectly related to our paper concern the effect of HFTs on price volatility, which is often (informally) considered to be an inverse measure of price efficiency (Carrion, 2013). In a study of several international equity markets, for instance, Boehmer et al. (2020) find that algorithmic trading intensity is positively correlated with price volatility. By contrast, Hasbrouck and Saar (2013) and Carrion (2013) find, for the NASDAQ, that HFTs dampen short term volatility. These contradicting results also provide impetus for further investigation into the effect of HFTs on the quality of asset prices.

2.2 HFTs and the single call auction

The revelation of information and price discovery in the call auction mechanism have been studied empirically in several previous studies. However, little has been said until now about the influence of HFTs on call auction prices.⁶ Our paper is part of a new strand of the microstructure literature that investigates the presence of HFTs in the call auction trading environment. [Bellia et al. \(2020, 2017\)](#) have recently investigated the role of HFTs in the preopening process of the Tokyo Stock Exchange and the Paris Stock Exchange. Although the detection of informed HFTs is not their focus, these two studies collectively provide evidence that HFTs drastically participate in the preopening phase and that they play a leading role in both the price formation process and the liquidity dynamics of the call auction market. A third study closely related to ours is that of [Boussetta et al. \(2020\)](#), who examine the overall role of the preopening phase in the cross-venue price discovery process. Using a set of intraday data from the Paris market (accompanied by a high-frequency trading indicator) and the BATS/Chi-X Europe multilateral trading facility, these authors find that opening prices in the Paris market contain information about the values of securities; although they do not conduct a direct test for the informativeness of HFTs' quotes, their findings are similar to ours. Additionally, these authors find that slow traders contribute the most to price discovery across the trading venues. Our empirical analysis complements the aforementioned studies in that we develop a direct method to test the existence of HFTs with accurate signals about the value of assets in the call market. Further, we assess the impact of HFTs' order placement activity on the informativeness of prices generated from the call auction mechanism.⁷

⁶[Ciccotello and Hatheway \(2000\)](#), [Cao et al. \(2000\)](#), [Barclay and Hendershott \(2003\)](#), [Barclay and Hendershott \(2008\)](#) and [Pagano et al. \(2013\)](#) investigate the NASDAQ dealers' market opening. [Madhavan and Panchapagesan \(2000\)](#) examine the NYSE opening procedure and the role of specialists in price discovery and price efficiency. In the order-driven trading framework, [Biais et al. \(1999\)](#), [Pagano and Schwartz \(2003\)](#), and [Hillion and Suominen \(2004\)](#) analyze the opening or closing auctions of the Euronext Paris stock market. [Comerton-Forde and Rydge \(2006\)](#) and [Moshirian et al. \(2012\)](#) study the opening and closing auctions in the Australian order-driven stock market. [Hauser et al. \(2012\)](#) examine the behavior of opening prices in the Tel Aviv Stock Exchange after the introduction of a random opening time, while [Kalay et al. \(2004\)](#) investigate, in the same market, the elasticity of (indicative) clearing prices during the preopening phase. Lastly, [Anagnostidis et al. \(2015\)](#) analyze the opening price discovery process in the Greek order-driven market. These empirical studies indicate, collectively, the importance of market stability in periods of increased price uncertainty, such as during the opening or closing of markets. Additionally, these studies suggest that the call auction is likely the most efficient known mechanism for revealing prices after periods of no trade.

⁷Note that the data set employed herein is identical to that used by [Bellia et al. \(2017\)](#) and part of that used by [Boussetta et al. \(2020\)](#).

2.3 Detecting informed trading

Our analysis is, also, related to a strand of the literature that concerns the detection of informed trading and its impact on prices. Microstructure theory predicts that the cost of trade should increase with the presence of information asymmetry (e.g., [Glosten and Milgrom, 1985](#)). In line with this theory, several studies use the bid–ask spread and/or the price impact of trades to quantify the level of information asymmetry ([Hasbrouck, 1991a,b](#); [Barclay and Warner, 1993](#); [Easley et al., 1996](#); [Huang and Stoll, 1997](#); [Chakravarty, 2001](#); [Anand and Subrahmanyam, 2008](#); [Menkhoff and Schmeling, 2010](#)). Bidding in the call auction framework, however, is only theoretical until the clearing time. From a technical perspective, therefore, these proxies of information asymmetry are less relevant during the order accumulation period. Moreover, recent empirical evidence suggests that inferences of information asymmetry may be biased when informed investors condition their trading on such measures of (il)liquidity ([Collin-Dufresne and Fos, 2015](#)). Other empirical studies assume *a priori* that certain groups of investors are informed. For example, [Boulatov et al. \(2012\)](#) directly employ the order flow of institutional investors as a proxy for informed trading.⁸ Such an approach risks sample selection biases; indeed, [Boulatov et al. \(2012\)](#) acknowledge that their findings do not support the argument that all institutions are informed.

Our paper differs from the studies mentioned above in several ways. Whereas existing studies focus on continuous trading, we attempt to detect informed HFTs in the call auction trading environment. Accordingly, we do not employ traditional spread–related or trade–related proxies for information asymmetry. Rather, our proposed test for detecting informed HFTs is based on the information content of crossing supply/demand schedules. Neither do we make any particular assumptions about the ability of HFTs (or non–HFTs) to infer true values (e.g., we do not *a priori* assume that HFTs hold information about future liquidation values). By doing so, we reduce the possibility of biased inference on HFTs’ quote informativeness when interpreting our empirical results.

⁸Because they invest in technologies and hire professionals with superior information processing skills, large institutions are frequently labelled as informed ([Hendershott et al., 2015](#)).

3 A test for detecting informed HFTs in the call market

3.1 The call auction framework

Following Kyle (1989), we consider an automated transparent one-period call market, in which a single risky asset with random ex-post liquidation value, $u \sim N(0, \sigma_u^2)$, with $\sigma_u^2 > 0$, and a risk-free asset are traded.⁹ We normalize the return of the risk-free asset to zero and, for the sake of simplicity, assume that market agents do not hold an initial endowment. The trading public consists of N informed, M uninformed and Z liquidity/noise traders ($N, M \in \mathbb{N}^0$ and $Z \in \mathbb{N}^+$).¹⁰ Informed and uninformed traders are assumed to behave rationally in the sense that, before posting their quotes, they update their expectations by observing the disseminated market price. After bidding, the market clears and trading takes place. Moreover, informed and uninformed traders are assumed to be risk averse and are endowed with constant absolute risk aversion (CARA) utility functions, $U(\pi_I) = -\exp(-\rho\pi_I)$ and $V(\pi_U) = -\exp(-\rho\pi_U)$, respectively, where ρ is the risk aversion coefficient, π_I is the terminal wealth for informed traders, and π_U is the terminal wealth for uninformed traders. Here, I denotes Informed and U denotes Uninformed. For simplicity, we assume that all agents (informed and uninformed) have the same risk aversion parameter (ρ).¹¹ Because they are risk averse, informed and uninformed traders submit limit orders, whereas liquidity/noise traders submit price-inelastic orders (e.g., market orders).

Informed investors hold a private signal (i.e., they evaluate the asset value, u , based on private information). This private signal, ι , with $\iota \sim N(0, \sigma_\iota^2)$, $\sigma_\iota^2 > 0$, corresponds to the sum of the ex-post liquidation value, u , and a white noise term, e , $\iota = u + e$, where $e \sim N(0, \sigma_e^2)$, with $e \perp u$ and $\sigma_e^2 > 0$. Uninformed traders do not hold any private information; they observe only the publicly available market price. Noise traders arrive in the market for liquidity purposes.

Informed investors, uninformed investors and noise traders have, respectively, the per capita demand functions X , Y , and Q , with $Q \sim N(0, \sigma_Q^2)$, $\sigma_Q^2 > 0$.¹² Informed investors' demand is linear with respect to their private signal, ι , and the public price, p : $X = \mu_I + \beta\iota - \gamma_I p$, with $\beta, \gamma_I > 0$. X implies that informed investors submit a downward sloping demand with respect to p , while they

⁹Without loss of generality, we assume that $E(u) = 0$ (e.g., Kovalenkov and Vives, 2014).

¹⁰We require that Z is strictly positive so that the Grossman and Stiglitz (1980) paradox is avoided.

¹¹Because there is no trading during the order accumulation period, we prefer to simplify the algebra by assuming that all agents (HFTs and non-HFTs) have the same level of risk aversion.

¹²We assume that the demand of noise trading is exogenous. Further, we assume that ι and Q are independent.

increase (decrease) their demand with bullish (bearish) information. Uninformed agents' demand is also downward sloping with respect to p : $Y = \mu_U - \gamma_U p$. Notice, in Y , that $\beta = 0$ inherently, as uninformed traders do not hold private information. The coefficients in the demand equations, X and Y , are determined by matching traders' expectations with their realizations, according to the rational expectations paradigm.

Both informed and uninformed traders act strategically; that is, they take into account the impact of their own orders on the clearing price. Formally, informed traders consider the effect of their demand, X , on the clearing price by observing the residual supply curve, p_I , such that $p = p_I + \lambda_I X$, with $\lambda_I > 0$. λ_I is the constant slope (i.e., the price elasticity), such that $\lambda_I = \partial p / \partial X$. Similarly, uninformed traders consider the effect of their orders on the market price by observing p_U , such that $p = p_U + \lambda_U Y$ with $\lambda_U = \partial p / \partial Y$ (see [Appendix A](#) for a detailed solution of the informed/uninformed traders' problem and [Appendix B](#) for the calculation of p_U , λ_U , p_I and λ_I).

The market clears at the price for which the excess demand is zero:

$$NX + MY + ZQ = 0. \quad (1)$$

Using equation (1), the market clearing price is calculated as follows:

$$p = \frac{1}{N\gamma_I + M\gamma_U} (N\beta\iota + ZQ + N\mu_I + M\mu_U), \quad (2)$$

with $\lambda = \frac{1}{N\gamma_I + M\gamma_U}$. Coefficient λ can be viewed as a measure of market liquidity (i.e., Kyle's λ).¹³

Theorem 5.1 in [Kyle \(1989\)](#) shows that, under certain conditions, there exists a unique symmetric linear equilibrium in such a trading framework. In particular, the equilibrium exists if $N \geq 2$ and $M \geq 1$, or $N \geq 3$ and $M = 0$, or $N = 0$ and $M \geq 3$. Also, an equilibrium exists if $N = 1$ and $M \geq 2$ while M is sufficiently large. By contrast, if $N + M \leq 2$ the equilibrium does not exist.

Price efficiency is measured with the amount of residual noise left after filtering the true price, u ,

¹³Notice that the market price is a linear combination of the private signal ι and the noisy liquidity demand Q . Indeed, equation (2) can be rewritten as $p = a\iota + bQ + c$ with $a = \frac{N\beta}{N\gamma_I + M\gamma_U}$, $b = \frac{Z}{N\gamma_I + M\gamma_U}$, and $c = \frac{N\mu_I + M\mu_U}{N\gamma_I + M\gamma_U}$.

with p (see [Appendix C](#) for the analytical calculation):

$$\text{Var}(u|p)^{-1} = \frac{1}{\sigma_u^2} + \frac{1}{\sigma_e^2 + \Delta^2 \sigma_Q^2}, \quad (3)$$

with

$$\Delta = \frac{Z}{N} \frac{1}{\frac{\lambda_I}{\delta} + \rho \sigma_e^2}, \quad (4)$$

where $\delta = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$. Equation (3) represents the informativeness of p about the liquidation value, u , of the asset.

[Table 1 about here]

Table 1 reports the behavior of price efficiency, $\text{Var}(u|p)^{-1}$, with respect to the exogenous model parameters. The model is implicit and thus the application of a numerical scheme is needed to obtain a solution. Price efficiency improves when the number of informed traders (N) increases, whereas it decreases with an increase in noise trading (Z). These results are natural, since informed traders' bids induce information shocks into the market price, whereas noise traders submit price-inelastic orders which decrease price efficiency. Also, an increase in the number of uninformed traders (M) flattens the residual supply curve against which informed traders optimize their trade. Thus, informed traders' bidding becomes less strategic (or else, more sensitive to the private signal) and, consequently, price efficiency improves. The risk aversion coefficient (ρ) is negatively correlated with price efficiency. Indeed, as traders become more risk averse, they reduce their aggressiveness and thus information is induced into prices at a lower rate. Lastly, notice that price efficiency increases as the precision of the private signal, $1/\sigma_e^2$, increases. A more detailed analysis of the limiting behavior of price efficiency with respect to the market parameters is provided in [Kyle \(1989\)](#).

3.2 HFTs in the call market

Before we proceed to incorporate high-frequency trading into the [Kyle \(1989\)](#) framework, it is useful to discuss the main characteristics of HFTs, as identified by the U.S. Securities Exchange Commission (SEC) and the European Securities and Markets Authority (ESMA). According to [SEC \(2010\)](#) and [Bouveret et al. \(2014\)](#), for ESMA, HFTs are characterized as proprietary traders who:

1. Use the co-location services and data feeds offered by the Exchanges, along with other electronic sources.
2. Use ultra high-speed algorithms for generating, routing, and executing orders.
3. Consider very short time-frames for establishing and closing positions.
4. Submit and cancel orders at extremely high frequencies.
5. Have as flat as possible positions at the end of the trading day.

It is evident that, because of the absence of actual trading during the order accumulation period in the call market, properties (2) through (5) become less relevant. By contrast, property (1) is of significant importance, as access to a broad set of public information may turn into an informational advantage during the pre-clearing phase (e.g., through the efficient collection and assessment of early morning and/or overnight news and quotes, during the preopening phase).¹⁴ Thus, for the remainder of our theoretical analysis, we use property (1) to distinguish HFTs from non-HFTs. That is, we shall assume that HFTs have direct access to broader sets of public information, as compared to non-HFTs (e.g., [Van Kervel, 2015](#); [Foucault et al., 2016](#)). This choice does not cancel out properties (2) through (5) which are equal in importance to property (1) when describing the activity of HFTs within the continuous trading framework. Moreover, property (1) does not guarantee that high-frequency trading algorithms interpret public information in a superior way, enabling a better assessment of the liquidation value of the asset than that implied by the current market price (i.e., in the context of our theoretical framework, property (1) does not guarantee that HFTs are informed).

The [Kyle \(1989\)](#) framework provides insight on the effect of the different types of traders (informed (N), uninformed (M), and noise (Z) traders) on the market price. We extend this framework by decomposing the trading public into HFTs and non-HFTs. More specifically, we assume that HFTs constitute a heterogeneous group of N_{HFT} informed, M_{HFT} uninformed, and Z_{HFT} noise traders, while non-HFTs constitute a second heterogeneous group of $N_{non-HFT}$ informed, $M_{non-HFT}$ uninformed, and $Z_{non-HFT}$ noise traders. By definition, we have the identities $N = N_{HFT} + N_{non-HFT}$, $M = M_{HFT} + M_{non-HFT}$ and $Z = Z_{non-HFT} + Z_{HFT}$. Notice that we do not make any *a priori* assumptions on the distribution of HFTs and non-HFTs (e.g., we do not *a priori* assume that

¹⁴For example, some segments in Xetra as well as the Bats/Chi-x multilateral trading facility, which inter-connects multiple European platforms, are also at the preopening phase before 09:00. Moreover, French firm earnings are announced around 08:00 ([Bellia et al., 2017](#)).

$N_{HFT} = 0$ or that $N_{HFT} > 0$). We do assume, however, that N and M adhere to the conditions derived by Kyle (1989) for the existence of a unique symmetric linear equilibrium (see the discussion in Section 3.1).¹⁵

3.3 Derivation of the test

After introducing HFTs into our analysis, we proceed to derive a test for detecting informed HFTs. Our first task is to obtain an estimate of the informativeness of HFTs' and non-HFTs' quotes. To do so, we calculate two theoretical prices, p_{HFT} and $p_{non-HFT}$, that are uniquely defined by the residual supply/demand schedules submitted by HFTs and non-HFTs, respectively. In other words, we seek for two theoretical prices, p_{HFT} and $p_{non-HFT}$, that minimize, respectively, the aggregate net demand formed by HFTs' orders and that of non-HFTs' orders, *ceteris paribus*.¹⁶ Figure 1 illustrates the idea of the construction of p_{HFT} from HFTs' aggregate net demand, for a hypothetical call market.¹⁷

[Figure 1 about here]

To calculate p_{HFT} , consider first the simple case where HFTs are only uninformed. Thus, $N_{HFT} = Z_{HFT} = 0$ and $M_{HFT} > 0$. By setting HFTs' aggregate net demand equal to zero ($N_{HFT}X + M_{HFT}Y + Z_{HFT}Q = 0$), we obtain:

$$p_{HFT} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 + \frac{Z^2 \sigma_Q^2}{N^2 \beta^2}} \left(\iota + \frac{ZQ}{N\beta} \right). \quad (5)$$

Using the projection theorem, we find that $Var(u|p_{HFT})^{-1} = Var(u|p)^{-1}$. That is, the efficiency of p_{HFT} is at the same level as market efficiency. This result is reasonable since all HFTs are uninformed and, therefore, they use the market price p to extract information about the liquidation value of the asset, u .

¹⁵Notice that we consider two types of informed traders (HFTs and non-HFTs) who observe the private signal, ι , with the same precision σ_ι . We could consider heterogeneous signals between informed HFTs and informed non-HFTs, in terms of precision, without changing the main predictions of the model.

¹⁶For instance, in the course of calculating p_{HFT} , each slope (computed according to the original optimization problems solved by informed and uninformed traders) is fixed at its equilibrium value. Put differently, to calculate p_{HFT} , we slide along the aggregate net demand curve, originally submitted by HFTs, until we find the price at which the aggregate net demand is zero (see also Figure 1). The same holds for the computation of $p_{non-HFT}$. It is important to note that uninformed and informed traders condition on p to solve their optimization problem, whereas p_{HFT} and $p_{non-HFT}$ are, ex-ante, unobserved variables.

¹⁷According to the assumptions of the model, p should be centered around zero. For illustration purposes, however, Figure 1 presents a positive p that is closer to reality. Because of symmetry, our analytic results hold in both cases.

Now, consider the more complex case where not all HFTs are uninformed, such that $N_{HFT} + Z_{HFT} > 0$ and $M_{HFT} \geq 0$. By setting HFTs' aggregate net demand equal to zero, we acquire

$$p_{HFT} = \frac{1}{N_{HFT}\gamma_I + M_{HFT}\gamma_U} (N_{HFT}\beta\iota + Z_{HFT}Q). \quad (6)$$

Using the projection theorem, the efficiency of p_{HFT} is given by

$$\text{Var}^{-1}(u|p_{HFT}) = \left(\sigma_u^2 - \frac{N_{HFT}^2\beta^2\sigma_u^4}{N_{HFT}^2\beta^2(\sigma_u^2 + \sigma_e^2) + Z_{HFT}^2\sigma_Q^2} \right)^{-1}. \quad (7)$$

Notice that $\text{Var}(u|p_{HFT})^{-1} \neq \text{Var}(u|p)^{-1}$. That is, price variations induced into p_{HFT} through informed and/or noise HFTs' orders, lead to deviations of p_{HFT} from p .

Symmetry allows us to follow the same steps when calculating $p_{non-HFT}$. Thus, if $N_{non-HFT} + Z_{non-HFT} = 0$ and $M_{non-HFT} > 0$, we obtain

$$p_{non-HFT} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 + \frac{Z^2\sigma_Q^2}{N^2\beta^2}} \left(\iota + \frac{ZQ}{N\beta} \right), \quad (8)$$

with $\text{Var}(u|p_{non-HFT})^{-1} = \text{Var}(u|p)^{-1}$. Further, if $N_{non-HFT} + Z_{non-HFT} > 0$ and $M_{non-HFT} \geq 0$, we get

$$p_{non-HFT} = \frac{1}{N_{non-HFT}\gamma_I + M_{non-HFT}\gamma_U} (N_{non-HFT}\beta\iota + Z_{non-HFT}Q), \quad (9)$$

with the following level of efficiency:

$$\text{Var}^{-1}(u|p_{non-HFT}) = \left(\sigma_u^2 - \frac{N_{non-HFT}^2\beta^2\sigma_u^4}{N_{non-HFT}^2\beta^2(\sigma_u^2 + \sigma_e^2) + Z_{non-HFT}^2\sigma_Q^2} \right)^{-1}. \quad (10)$$

Notice that, as in the case of p_{HFT} , $\text{Var}(u|p_{non-HFT})^{-1} \neq \text{Var}(u|p)^{-1}$.

Having calculated the information content of HFTs' and non-HFTs' quotes, we now turn to an extensive comparative analysis. We compare the information content of p , p_{HFT} , and $p_{non-HFT}$, by systematically exploiting the relations between $\text{Var}^{-1}(u|p)$, $\text{Var}^{-1}(u|p_{HFT})$, and $\text{Var}^{-1}(u|p_{non-HFT})$. Our results are summarized in the following proposition which provides sufficient and necessary

conditions for all of the different possibilities (see [Appendix D](#) for the proof).

Proposition 3.1. *For $Var^{-1}(u|p)$, $Var^{-1}(u|p_{HFT})$, and $Var^{-1}(u|p_{non-HFT})$, the following relations hold:*

1. *If $N_{HFT} + Z_{HFT} > 0$ and $N_{non-HFT} + Z_{non-HFT} > 0$, then*

$$Var^{-1}(u|p) \geq Var^{-1}(u|p_{HFT}) \quad \Leftrightarrow \quad \frac{Z_{HFT}}{N_{HFT}} \geq \frac{Z}{N} \quad (11)$$

$$Var^{-1}(u|p) \geq Var^{-1}(u|p_{non-HFT}) \quad \Leftrightarrow \quad \frac{Z_{non-HFT}}{N_{non-HFT}} \geq \frac{Z}{N} \quad (12)$$

$$Var^{-1}(u|p_{HFT}) \geq Var^{-1}(u|p_{non-HFT}) \quad \Leftrightarrow \quad \frac{Z_{non-HFT}}{N_{non-HFT}} \geq \frac{Z_{HFT}}{N_{HFT}} \quad (13)$$

2. *If $N_{HFT} = Z_{HFT} = 0$, $M_{HFT} > 0$ and $N_{non-HFT} + Z_{non-HFT} > 0$, then*

$$Var^{-1}(u|p_{non-HFT}) = Var(u|p)^{-1} = Var^{-1}(u|p_{HFT}) \quad (14)$$

3. *If $N_{HFT} + Z_{HFT} > 0$, $N_{non-HFT} = Z_{non-HFT} = 0$ and $M_{non-HFT} > 0$, then*

$$Var^{-1}(u|p_{non-HFT}) = Var(u|p)^{-1} = Var^{-1}(u|p_{HFT}). \quad (15)$$

Notice that the comparison of the three levels of price efficiency in proposition (3.1) depends exclusively on the amount of noise trading ($Z_{HFT}, Z_{non-HFT}, Z$) and the amount of informed trading ($N_{HFT}, N_{non-HFT}, N$). For each comparison pair (e.g., HFTs versus non-HFTs), the ratio of the number of noise traders over the number of informed traders determines which price is the more efficient. This is an intuitive result, since uninformed traders act as price learners, whereas informed and noise traders' order flow induces exogenous shocks to the supply/demand schedule. Note also that the magnitude of strategic trading ($\lambda_{I,U}$), as well as other exogenous parameters, such as the precision of informed traders' signal ($1/\sigma_e^2$) or the number of uninformed traders ($M, M_{HFT}, M_{non-HFT}$), are absent from the comparison formulas in proposition (3.1).¹⁸

The relations of proposition (3.1) can be further exploited to derive a theoretical test for the

¹⁸An interesting corollary concerns the extreme case where HFTs consist of only informed or of only noise traders, which translates into $N_{HFT} > 0, Z_{HFT} = M_{HFT} = 0$ or $Z_{HFT} > 0, N_{HFT} = M_{HFT} = 0$, correspondingly. Here, the

detection of informed HFTs, which is our final task. This test is summarized in the following proposition (see [Appendix D](#) for the proof).

Proposition 3.2. *If one of the following conditions is met, then $N_{HFT} > 0$; i.e., the group of HFTs comprises a subgroup of informed traders:*

1. $Var^{-1}(u|p_{non-HFT}) < Var^{-1}(u|p_{HFT})$
2. $Var^{-1}(u|p) < Var^{-1}(u|p_{HFT})$.

Conditions (1) and (2) in proposition (3.2) are designed to detect informed HFTs by capturing deviations of the information content of p_{HFT} from the information content of the price derived from the non-HFTs' aggregate net demand, $p_{non-HFT}$ (condition 1), or from the information content of the market price p (condition 2; see, also, [Figure 1](#)). Given a data set with a high-frequency trading classification, proposition (3.2) can be directly utilized to empirically test the presence of informed HFTs in the call market. Moreover, this proposed empirical test has the advantage of not requiring an estimate of $Z, N, Z_{HFT}, N_{HFT}, Z_{non-HFT}$, and $N_{non-HFT}$, which are, in general, unknown numbers of market participants. As a last comment, we note that, although satisfaction of conditions (1) and (2) conclusively asserts the existence of informed HFTs, non-satisfaction does not imply their absence.

4 Empirical analysis

4.1 The Paris Euronext market

Trading on the Euronext Paris platform is conducted in two main ways: the order-driven market model and the liquidity provider (LP) quote-driven model. In the former, liquidity is provided through individual brokers' orders, along with the enhancement of supplementary liquidity providers (SLPs); in the latter, liquidity is exclusively provided via the quotation of designated liquidity providers (LPs).

efficiency $Var^{-1}(u|p_{HFT})$ becomes constant and equal to $(\sigma_u^2 + \sigma_e^2)/(\sigma_u^2\sigma_e^2)$ or $1/\sigma_u^2$, respectively, yielding

$$\frac{1}{\sigma_u^2} < Var^{-1}(u|p) < \frac{\sigma_u^2 + \sigma_e^2}{\sigma_u^2\sigma_e^2}.$$

By symmetry, the same corollary applies to non-HFTs as well.

The order-driven system includes either continuous or periodic auction trading. The first mechanism is concerned with the more liquid securities, like those comprising the CAC 40 Index, while the second is for the less liquid securities. The continuous system, examined herein, is operated under the following daily time schedule (in Central European Time):¹⁹

1. 07:15–09:00 Preopening phase – order accumulation period
2. 09:00 Opening auction
3. 09:00–17:30 Main session – continuous trading
4. 17:30–17:35 Preclosing phase – order accumulation period
5. 17:35 Closing auction
6. 17:35–17:40 Trading at the last phase (at the close)
7. 17:40–07:15 After hours trading

The opening call auction procedure lasts 1 hour and 45 minutes. During this time period investors are allowed to submit, modify, or cancel orders, while observing the disclosed information on the evolution of the indicative clearing price–volume pair. Since there is no actual trading before the official opening of the market, all orders are stored into the central limit order book with price–time execution priority. Three main types of orders are allowed during the preopening phase: a) market on opening orders, b) pure market orders, and c) limit orders.²⁰ Market on opening orders and pure market orders have priority over limit orders at the time of the auction. Figure 2 illustrates the formulation of crossing supply and demand lines, due to the absence of trade, for a hypothetical set of limit prices and quantities. Notice that, because stock prices are discrete, it is possible for more than one equilibrium value to be present.

[Figure 2 about here]

At the end of the accumulation period, the electronic system considers the supply–demand schedule formed by the queuing orders, seeking the price that maximizes the trading volume. If the maximum volume principle suggests more than one equilibrium price, the opening price is set according

¹⁹The trading day schedule can be found at: <https://www.euronext.com/en/trading-calendars-hours>

²⁰Because they do not include price preference, market and on open sell (buy) orders are aggregated at the best ask (bid).

to the minimum volume surplus principle. Lastly, if more than one price satisfies the minimum surplus principle, then the system fixes as the opening price the one closest to the reference price; this reference price is the price of the last trade on the central order book before the preopening phase. After the opening price is set, buy and sell orders are matched and executed in a single trade and at a single opening price. Unexecuted market or limit orders are sent on to the main session with the original price and time priority, with market orders being stored as limit orders at the opening price.

It is useful to note that the opening time in the Paris market was fixed at 09:00 until 15/08/2015, whereas after 15/08/2015 the uncrossing takes place randomly between 09:00:00 and 09:00:30.

4.2 *Data sample*

The data used in the present study are extracted from the AMF–BEDOFIH Paris Euronext high–frequency database. Our sample includes 36 stocks from the CAC 40 Index composition on January 3, 2013, for the year 2013 (254 trading days).²¹ For the remaining four stocks in the index, transactions take place on the Paris platform, while order placement is conducted on other Euronext platforms (see Table 3). We have excluded these stocks from our sample because the database does not contain information on the related quoting activity.

The sample comprises two main files: i) trades and ii) orders. The first file contains information about the trading history in the Paris market: time (accurate to the microsecond), price, and traded quantity. The second file contains information about the order placement history: time of submission, price, size, side, duration, type, validity, and time of release from the system (either because of execution or because of cancellation).

In our sample period (2013), the opening time is fixed at 09:00 (see, also, Section 4.1). This feature allows for investors to adjust (or cancel) their orders just before the opening, without fearing of having them executed at an undesirable clearing price, as opposed to auction mechanisms with a random clearing time. Moreover, informed investors may be motivated to delay their bidding until

²¹The Paris market data included in the BEDOFIH database are provided by the AMF. Further Information on the BEDOFIH database can be found at: <https://www.eurofidai.org/en>.

seconds before the clearing time, in order to preserve their informational advantage.^{22,23}

An aspect of the data set crucial to our work is a flag that indicates whether an order (limit or market) is submitted by an HFT (Megarbane et al., 2017). The high-frequency trading flag categorizes the trading public as follows:

- a) Mixed high-frequency traders (hereafter, MIXED): large institutions (for instance, investment banks) applying high-frequency trading algorithms.
- b) Other high-frequency traders (hereafter, OHFTs): other firms applying high-frequency trading algorithms.
- c) Non-high-frequency traders (hereafter, non-HFTs).

The high-frequency trading flag, provided by the Autorité des Marchés Financiers (AMF), is based on a market participant meeting one of the following two conditions:

- 1) A market participant (with a specific ID) is classified as an HFT if: i) the average lifetime of his/her canceled orders is less than the average lifetime of all canceled orders in the book, and ii) he/she has canceled at least 100,000 orders during the year.
- 2) A market participant (with a specific ID) is classified as an HFT if: i) he/she has canceled at least 500,000 orders with a lifetime of less than 0.1 second, and ii) the top percentile of the lifetime of his/her canceled orders is less than 500 microseconds.

If, additionally, the market participant operates on behalf of a large investment bank, he/she is classified as a MIXED trader. Further, once a trader has been classified as MIXED, OHFT, or non-HFT, this flag is immutable.²⁴

²²See, for example, Medrano and Vives (2001) for a detailed analysis of the effect of a random opening time on investors' order placement strategies. Note that we cannot exclude the possibility that some form of price manipulation could take place during the order accumulation period. However, a limitation of the data set employed in this study is that it does not include traders' IDs and, therefore, we cannot accurately track traders' submit/cancel activities. Thus, it is not possible to directly test the price manipulation hypothesis.

²³For additional information on the Paris stock market opening and main session trading, see, also, the seminal works by Biais et al. (1999) and Biais et al. (1995), respectively.

²⁴It is important to mention that although large institutions typically involve both HFT and non-HFT operations (e.g., OTC trading), MIXED messages represent the HFT activities of these institutions. Also, messages flagged as MIXED may represent the quoting activity of traders who operate on their own account (i.e., proprietary traders) or on behalf of clients. The information content of preopening quotes submitted by these sub-groups of MIXED traders is beyond the scope of the current study. For more details on this issue, see the analyses in Boussetta et al. (2020) and Bellia et al. (2017).

The AMF definition is in line with the description, quoted in Section 3.2, of property (4) of high-frequency trading (i.e., HFTs submit and cancel orders at extremely high frequencies), as defined by the SEC and ESMA. By contrast, our theoretical framework considers property (1) (i.e., the use of co-location services and data feeds offered by exchanges, along with other electronic sources) when separating HFTs from non-HFTs. Since the AMF database does not include information about HFTs' access to multiple data sources, however, we will assume that the SEC and ESMA's property (1) applies to the MIXED HFTs and Other HFTs flagged by AMF.

4.3 Preliminary analysis

4.3.1 Order flow during the preopening

To further substantiate the empirical methodology, this section provides a preliminary analysis of the preopening order flow and the opening trading activity in the Paris market.

[Table 2 about here]

Table 2 reports summary statistics on submissions, modifications, and cancellations during the preopening phase in 15-minute intervals for each trader-type: MIXED traders, OHFTs, and non-HFTs. As can be seen, the order placement activity of MIXED traders and OHFTs rapidly increases after 08:30. For OHFTs, the (across stocks and days) average number of submissions in the intervals 08:30–08:45 and 08:45–09:00 (opening) is, respectively, 220.8 and 112.6, while before 08:30 it is practically equal to zero. Similarly, for MIXED traders, the level of submissions increases between 08:45 and 09:00, the corresponding stock-day average number of submissions being 129.9, while before 08:45 submissions are limited. By contrast, non-HFTs enter the market as soon as the preopening process begins (i.e., at 07:15). Along with MIXED traders and OHFTs, non-HFTs increase their bidding activity after 08:30.

Regarding modifications and cancellations, MIXED traders and OHFTs sharply increase their activity just before the opening. In the interval 08:45–09:00, OHFTs submit, on average, 36.6 modifications and 30.6 cancellations. For MIXED traders, the corresponding values are 58.9 and 33.3. By contrast, the number of modifications or cancellations is close to zero for both MIXED traders and OHFTs, before 08:45. non-HFTs also increase their modifications and cancellations a

few minutes before the opening (submitting, on average, 17.5 and 6, respectively). However, their order adjustment activity is limited compared to that of MIXED traders and OHFTs.

Interestingly, an increase in the average order size is observed for all types of traders before the opening. On average, OHFTs increase their order size from 192.1 to 667.9 shares, between 08:30 and 09:00. In the same interval, MIXED traders and non-HFTs increase their bidding intensity from 1,824.3 to 5,938.6 shares and from 779.8 to 1,373.8 shares, respectively. These figures suggest, also, that the most aggressive orders, in terms of size, are submitted by MIXED traders just prior to opening.

It is important to note that, according to our computations, the percentage of quotes placed by designated supplementary liquidity providers during the preopening phase is less than 2%, which is practically negligible. Hence, we postulate that the rest of the traders (MIXED, OHFTs, and/or non-HFTs) entering the market before the clearing time are willing to trade for individual and not for market making purposes.

Given that our overall purpose is to investigate the informativeness of HFTs' quotes, for the remainder of our empirical analysis we will focus on the last 30 minutes of the preopening phase (08:30–09:00), during which time HFTs rapidly enter the market.²⁵

4.3.2 The last 30 minutes

In this section, focusing on the last 30 minutes of the preopening phase, we provide visual evidence of the order flow and market depth characteristics attributed to the three types of traders in our data set. Figure 3 illustrates the evolution of the across days and stocks average number of submissions, modifications and cancellations for each type of trader (MIXED traders, OHFTs, and non-HFTs) in 1-minute time intervals. OHFTs rapidly enter the market around 08:30; the average number of submissions is almost equal to 140 in the interval 08:30–08:31. For non-HFTs, the number of submissions is also increased in the interval 08:30–08:31, whereas MIXED traders exhibit a very low submission rate at that point. OHFTs' order submission frequency is also notably higher between 08:45 and 08:46.

[Figure 3 about here]

²⁵When we conduct our empirical analysis over the entire preopening phase, our results show that before 08:30 indicative prices are highly noisy.

Inspecting the details of the order submission activity in the intervals 08:30–08:31 and 08:45–08:46, we find the following. i) A large portion of order submissions within these intervals is followed by a “valid for auction” instruction; thus they were never meant to be executed in the continuous session. Unless these orders are canceled by the broker, they are either executed in the opening call or, if not, are forwarded to the closing call (which we find happens rather frequently). ii) The majority of these orders exhibit unique characteristics, which distinguish them from the rest of the order placement activity. In particular, these are limit orders (i.e., they have a price preference) submitted sequentially in time. For instance, we often observe a group of sequentially submitted buy orders followed by a group of sequentially submitted sell orders. More important, these buy orders decrease in price, whereas the sell orders increase in price. iii) The best bid and the best ask, in such a sequence of orders, usually define a positive spread (as in the continuous session). iv) Frequently, the previous closing price is within the spread; that is, below the best ask and above the best bid. v) Such order submission sequences, observed in several of our sample stock–days, are mainly placed by OHFTs (they may possibly have been submitted by a single OHFT). Nonetheless, we have also found such patterns stemming from non–HFTs.

[Figure 4 about here]

Because these order patterns are randomly submitted within the intervals 08:30–08:31 and/or 08:45–08:46, it is hard to estimate the across days and stocks average frequency of their occurrence. Thus, we mostly rely on a visual inspection of the order files. To illustrate, Figure 4 plots a supply/demand schedule submitted on 14/03/2013 between 08:45:48 and 08:45:50 (i.e., within 2 seconds) for a particular stock (Credit Agricole). All orders are sequentially submitted, meaning that there are no additional messages interrupting the order submission pattern. The trader type is OHFT, the best ask is 7.2 Euros, the best bid is 7.16 Euros, and the previous close is 7.186 Euros. Further, all orders are classified as “valid for auction.” A possible interpretation for such a sequence of orders is that they are submitted by a single algorithm that attempts to profit from a large price swing at the opening, with respect to the previous close. This order sequence does not imply a direction toward an informative signal; rather, the OHFT engages in both a long and a short position. We argue that such orders should convey little information about the asset’s liquidation value, adding noise to the price formation process.

After 08:55, that is, 5 minutes before the opening, all types of traders increase their order submission activity (Figure 3). The same holds for order modifications and cancellations, especially for MIXED traders and OHFTs. We postulate that these HFTs increase their cancellations and modifications a few minutes before the clearing time, to resubmit larger and more aggressive orders with the intention of having them executed. Thus, orders made 5 minutes or less before the opening should convey information about the stock’s liquidation value (Biais et al., 1999).

[Figure 5 about here]

Figure 5 depicts the evolution of the market depth attributed to MIXED traders’, OHFTs’, and non-HFTs’ shares during the last 30 minutes of the order accumulation process. The top graph illustrates the total number of submitted shares (buy and sell side), averaged across days and stocks. non-HFTs’ shares constitute the vast majority until 8:56, while after 08:56 MIXED traders’ liquidity provision is prevalent. For OHFTs, the average number of submitted shares slightly increases throughout the preopening phase. These figures are consistent with the results presented in Table 2, according to which the average number of submissions, as well as the average order size, increases during the last minutes of the preopening phase for MIXED traders and OHFTs.

The bottom graph of Figure 5 depicts the total number of shares (both to buy and to sell) submitted by MIXED traders, OHFTs, and non-HFTs, on the executable side of the book. These are the shares that determine the clearing price/volume pair since the call auction algorithm searches for the price that maximizes (minimizes) the clearing volume (excess demand).²⁶ Additionally, we plot the evolution of the indicative clearing volume multiplied by two, as trading involves both the buy and the sell sides. This multiplication thus allows us to compare the clearing volume and the trader-type volume paths (which sum the available buy side and sell side shares). The clearing volume increases almost exponentially toward the opening, as does the volume submitted by MIXED traders and non-HFTs. These figures suggest that a significant number of the shares submitted by MIXED traders and non-HFTs, prior to opening, are serious orders meant to be executed. For OHFTs, there is also an increase in order aggressiveness prior to opening. This increase, however, is smaller than that observed for MIXED traders and non-HFTs.

²⁶Given that the indicative/clearing price is p , executable buy (sell) orders are those with a limit price higher than or equal to (lower than or equal to) p . Also, market and at-the-open orders with no price preference are considered to be infinitely aggressive and thus they have the highest execution priority.

4.3.3 Opening trade

This section discusses summary statistics at the opening of the market. Table 3 reports the across days means for the following variables: opening volume, number of trades, and value of transactions. Additionally, we present the value of the opening transaction as a percentage of the total daily transaction value, as well as the percentages of MIXED traders', OHFTs', and non-HFTs' activity on the buy and the sell sides of the market. The opening trade accounts for approximately 1.4% of the total daily traded volume. This percentage is rather low, compared to the 10% reported by [Biais et al. \(1999\)](#) for the opening auction of the CAC 40 stocks in 1993. It is consistent, however, with our theoretical analysis, in which we have chosen to model the clearing market using a finite number of agents.

In terms of high-frequency trading, the percentage of MIXED flagged trades at the opening is substantially increased; almost 46.3% of trading on the buy side and 40.5% on the sell side involves MIXED traders (on average, across stocks). By contrast, only 5.2% (4.5%) of buy (sell) side trading involves OHFTs' orders. This result suggests that the vast majority of orders submitted by OHFTs in the intervals 8:30–8:31 and 8:45–8:46 are forwarded to the next auction (i.e., the closing call). That is, the opening price is not volatile enough, on average, to trigger the execution of these orders (see, also, Section 4.3.2). Lastly, a significant amount of non-HFTs' orders are executed at the open (almost 48.5% on the buy side and 55% on the sell side). Overall, these results indicate that MIXED traders' and non-HFTs' orders may contain valuable information about the liquidation value of the asset at the opening trade.

[Table 3 about here]

4.4 Empirical methodology and results

Having demonstrated that HFTs participate actively in the preopening phase and in the opening trade, we next address our main research question: Are HFTs' quotes informative?

4.4.1 Empirical prices

To apply the empirical test described by proposition (3.2), we first need to calculate p , p_{HFT} , and $p_{non-HFT}$. To do so, we sort the messages in our historical database into the following three sets:

1. The full order book: the set of all quotes submitted in the system.

2. The HFTs' order set: the set of orders submitted by MIXED traders and OHFTs, directly pertaining to high-frequency trading activity. To remain consistent with our theoretical framework, we merge the MIXED traders and OHFTs into one group, called the HFTs group. Thus, MIXED traders and OHFTs may belong to Z_{HFT} , N_{HFT} and/or M_{HFT} . Some OHFTs, for example, may be noise traders (i.e., belonging to Z_{HFT}), who submit uninformative orders in the intervals 08:30-08:31 and 08:45-08:46. On the other hand, some MIXED traders, representing large institutions, may be better informed than the other market participants (i.e., they may comprise a subgroup of N_{HFT}).
3. The non-HFTs' order set: the set of slow traders' quotes.

For each of the three sets, we calculate the indicative price, at which the aggregate net demand is minimized, in 1-minute intervals from 08:30 to 09:00 (opening). According to our theoretical framework, the aggregate net demand should be equal to zero. However, because of price discreteness (see Figure 2), the opening rules set a clearing price at which the clearing volume is maximized, or in other words, at which the aggregate net demand is minimized. We denote these three empirical prices as $p_{FULLBOOK}$, p_{HFT} , and $p_{non-HFT}$, and use them to conduct our empirical test.^{27,28}

4.4.2 Econometric methodology

To conduct our empirical test, we need to acquire an estimate of the informativeness of $p_{FULLBOOK}$, p_{HFT} , and $p_{non-HFT}$. We employ the unbiasedness regression technique of Biais et al. (1999), taking the current day closing price as a representation of the value of the stock, u . This proxy is based on the assumption that at the end of the trading day prices reflect all market information.²⁹ The previous day's closing price represents the expected value of the stock at time zero ($t = 0$) conditional on I_0 , $E(u|I_0)$, with I_0 being the information set at the start of the preopening phase (i.e.,

²⁷When the market is empty (in other words, when supply does not cross demand) we fix the mid-point value as the indicative price (Hasbrouck, 1991a). This feature is observed rather often for OHFTs before 08:30, as their participation in the market is practically negligible. When there are no orders on the buy and/or the sell side we omit the calculation.

²⁸In a similar way, Madhavan and Panchapagesan (2000) and Anagnostidis et al. (2015) separate market from limit orders in the NYSE and the Athens Exchange market opening, respectively, to compare the information content of the actual system price (which includes price-inelastic market and at-the-open orders) with the information content of the limit order book.

²⁹We have also used the current day opening price and the current mid-day price as proxies of the liquidation value of the stock; these results, available upon request, are qualitatively very similar to those reported.

the start of the trading day). If the indicative price at time $t > 0$, p_t , reflects information about u , then $p_t = E(u|I_t)$, with I_t being the information set at time t . On the other hand, if the price, p_t , incorporates noise, then it should reflect the information set I_0 plus a noise term, η_t . Thus, $p_t = E(u|I_0) + \eta_t$, with $\eta_t \perp u$.

To operationalize these arguments, we use the following overnight logarithmic returns: 1) the previous day's close to the current day's preopening time, t , R_{ct} ; and 2) the previous day's close to the current day's close, R_{cc} . R_{ct} proxies the difference between the preopening price, p_t , and the equilibrium value at the start of the day, $E(u|I_0)$: $[p_t - E(u|I_0)]$. R_{cc} proxies the change of the equilibrium price of the stock, $[u - E(u|I_0)]$, at the end of the trading day. To infer price efficiency, we examine the correlation between the two types of returns over the sample trading days, at each preopening time stamp.³⁰ Specifically, for each security, we estimate the following linear regression:

$$R_{cc} = a + bR_{ct} + \epsilon, \quad (16)$$

where t is the preopening time stamp, using the three prices, $p_{FULLBOOK}$, p_{HFT} and $p_{non-HFT}$, in the calculation of R_{ct} (summary statistics for the calculated R_{cc} and R_{ct} returns are presented in [Appendix E](#)).

After we estimate equation (16), two statistics allow us to infer price informativeness. The first is the slope coefficient, b . If b is not significantly different from unity, then the stock's tentative preopening prices are efficient, as they reflect all the market information available at time t (I_t). By contrast, if b is different from unity, then the indicative prices reflect information plus some noise.³¹ In other words, the estimated b coefficient can be viewed as a signal to noise ratio ([Barclay and Hendershott, 2003](#)). The second statistic, and the one more relevant to our analysis, is the variance of the residual term in equation (16), ϵ , which measures the remaining uncertainty about the value of the stock. The inverse of this variance constitutes a measure of price efficiency (that is, a direct

³⁰[Boussetta et al. \(2020\)](#) also apply the unbiasedness regression technique to infer price efficiency in the Paris market. Our approach, however, differs in that we construct separate prices on the basis of HFTs' and non-HFTs' quotes.

³¹Note that stock price efficiency at the opening can also be examined using variance ratio statistics calculated on the basis of daily data (for instance, opening and closing prices) ([Stoll and Whaley, 1990](#)). The unbiasedness regression technique, however, has the advantage of allowing us to examine the dynamics of stock prices within the trading day. Further, by running a separate regression for each consecutive time interval within the preopening phase, we avoid the nonstationarity issues that arise due to price adjustments as the price discovery evolves.

estimate of the statistic $Var(u|p)^{-1}$, allowing us to use proposition (3.2) to infer the presence of informed HFTs in the market. Following [Biais et al. \(1999\)](#), we calculate the root mean square error ($RMSE$) of the regression and use its inverse ($RMSE^{-1}$) as an estimate of price informativeness.

[Figure 6 about here]

4.4.3 The information content of quotes

The upper graph in Figure 6 plots the average estimated b coefficients, across stocks, in 1-minute intervals during the last 30 minutes of the preopening phase for the three cases: $p_{FULLBOOK}$, p_{HFT} , and $p_{non-HFT}$. For $p_{FULLBOOK}$, it is evident that prices gradually incorporate information until they become fully efficient at the opening. According to the reported confidence intervals at the opening, the b coefficient is not significantly different from unity at 09:00.³² Thus, there is evidence of a “learning” process, in which information is gradually revealed toward the opening. This learning pattern is similar to that reported by [Biais et al. \(1999\)](#) and [Boussetta et al. \(2020\)](#) for the Paris market, by [Ciccotello and Hatheway \(2000\)](#) and [Barclay and Hendershott \(2003\)](#) for the NASDAQ, by [Comerton-Forde and Rydge \(2006\)](#) and [Moshirian et al. \(2012\)](#) for the Australian Exchange, and by [Anagnostidis et al. \(2015\)](#) for the Athens Exchange.³³ On the other hand, the early prices are rather noisy, reflecting the fact that the bidding activity around 08:30 (see Section 4.3) does not convey information about the liquidation value.

Looking at p_{HFT} , we can see that the HFTs’ early orders contaminate the price formation process with large amounts of noise. The b coefficient is less than 0.2 at 08:30 (i.e., when OHFTs’ rapidly submit orders into the system). Closer to opening, HFTs increase their order aggressiveness, driving the price toward its true value. The information content of p_{HFT} increases rapidly within the last minutes of the preopening phase, as b approaches unity. At the opening, however, b is slightly less than 0.8 and statistically different from unity at the 5% probability level, suggesting that p_{HFT}

³²[Barclay and Hendershott \(2003\)](#) point out that if stock returns are serially uncorrelated and measured without any errors, then the slope coefficient b in equation (16) should equal unity. Otherwise, because of the presence of various microstructure effects that induce correlation into returns (e.g., temporary pricing errors and the non-synchronous trading effect), b should become noisy, deviating from unity. As in [Barclay and Hendershott \(2003\)](#), in the present study we report confidence intervals using the time-series standard errors from the coefficient estimates. We have also replicated the regression analysis using Newey–West standard errors corrected for serial correlation and heteroscedasticity. Results, available upon request for the interested reader, are similar with those reported herein.

³³[Boussetta et al. \(2020\)](#) use the SBF 120 Index constituents for years 2012 and 2013, whereas we use the CAC 40 Index (for year 2013) which includes the most liquid securities in the French market.

still reflects some noise. For $p_{non-HFT}$, the b coefficient increases monotonically with time, hinting that non-HFTs' quotes do contain information. However, b remains rather low during the entire 30 minute period, ranging approximately from 0.2 to 0.45.

It is interesting to notice the small drop of b at 08:50 in the p_{HFT} case, followed by an increase prior to opening. To further explore this feature, in Figure 7 we plot the average, across days and stocks, 1-minute logarithmic return, Δp_k , $k = 1, \dots, 29$, during the last 30 minutes of the preopening phase for the three prices: $p_{FULLBOOK}$, p_{HFT} , and $p_{non-HFT}$. In addition, to better understand the magnitude of price variability during the last 30 minutes of the preopening phase, we plot the average, across days and stocks, total price change from 08:30 to 09:00, $\Delta p^{FULLBOOK}$, calculated using the indicative price $p_{FULLBOOK}$ obtained from the full book (horizontal line). Notice the peak in p_{HFT} returns at 08:50, although prices later reverse to their mean, prior to opening. This pattern is likely related to the temporary market stress at 08:50 due to the opening procedure of the Frankfurt Exchange. Xetra disseminates the first indicative auction equity prices at 08:50, while several Eurex derivatives (e.g., the EURO STOXX 50[®] Index Options) enter the continuous trading phase at 08:50. Note, however, that in the present study we do not formally test for spillover effects between the two exchanges.

[Figure 7 about here]

[Table 4 about here]

We next calculate the root mean square error ($RMSE$) of the regression which we use to reveal the presence of informed HFTs in the preopening phase. The bottom graph in Figure 6 plots the inverse of $RMSE$, which is the estimate of price efficiency in the context of our theoretical framework, from 08:30 to 09:00. Table 4 summarizes the results of our comparison of price efficiency for p , p_{HFT} , and $p_{non-HFT}$ from 08:30 to 09:00 in 1-minute frequency. We compare price efficiency by calculating the following differences: i) $\Delta FH = RMSE_{FULLBOOK}^{-1} - RMSE_{HFT}^{-1}$, ii) $\Delta FN = RMSE_{FULLBOOK}^{-1} - RMSE_{non-HFT}^{-1}$ and iii) $\Delta HN = RMSE_{HFT}^{-1} - RMSE_{non-HFT}^{-1}$. For each case, we report the across stocks percentage of positive differences, the across stocks average difference, and the t-statistic obtained from a t-test for the null hypothesis that the across stocks average difference is not significantly different from zero at the 5% probability level.

All prices exhibit equal price efficiency early in the preopening phase, from 08:30 to 08:37, with all t -statistics, as reported in Table 4, being less than 1.96 in absolute value. In addition, the low levels of the inverse of the $RMSE$, depicted in Figure 6, suggest that the early preopening orders are rather noisy, which is consistent with the preopening paths of the b coefficient estimates, as illustrated in the upper graph of Figure 6. In view of equation (13) in proposition (3.1), these findings suggest that the ratio of noise to informed trading is equally high within the subgroups of HFTs and non-HFTs. That is, on average, the ratios Z_{HFT}/N_{HFT} and $Z_{non-HFT}/N_{non-HFT}$ should be approximately equal, while $Z_{HFT} \gg N_{HFT}$ and $Z_{non-HFT} \gg N_{non-HFT}$. Notice, however, that neither condition (1) nor (2) of proposition (3.2) is met; therefore, we are not able to derive a conclusion about the presence of informed HFTs (i.e., that $N_{HFT} > 0$).

After 08:37, significant price variations emerge. The indicative price, $p_{FULLBOOK}$, becomes more efficient than $p_{non-HFT}$. The average difference ΔFN is positive and statistically significant at the 5% probability level until the opening. Also, after 08:50 and until the opening, $p_{FULLBOOK}$ is more informative than p_{HFT} , as the average difference, ΔFH , is positive and statistically significant at the 5% level (Table 4). Importantly, after 8:56, four minutes before the opening time, p_{HFT} becomes more informative than $p_{non-HFT}$. The average difference, ΔHN , is increasingly positive and statistically significant at the 5% level until the opening (Table 4). Additionally, the across stocks percentage of positive differences, ΔHN , increases from 58.33% to 97.22%. In view of proposition (3.2), these results indicate that after 08:56 informed HFTs are present in the market (i.e., $N_{HFT} > 0$). Accordingly, in line with proposition (3.1), the fact that $\Delta HN > 0$ implies that the ratio of noise to informed trading is significantly lower within the subgroup of HFTs as compared to the same ratio within the subgroup of non-HFTs; that is, $Z_{HFT}/N_{HFT} < Z_{non-HFT}/N_{non-HFT}$. These findings coincide with the HFTs' large and aggressive (in size and price) order submission activity, particularly the activity of MIXED traders who represent large institutions, prior to opening. Overall, our evidence suggests that HFTs' quotes are highly informative prior to opening (and more informative than non-HFTs' quotes), driving the indicative price to the level of the fair value of the stock.³⁴

³⁴This result is consistent with the evidence provided by Bellia et al. (2017). Using the BEDOFIH CAC 40 sample for year 2013, these authors find that HFTs in the Paris opening should profit from orders submitted in the last minutes of the preopening, by assuming that the liquidation value of the stock is represented by the main session price a minute after the opening. Based on this finding, the authors indirectly conclude that HFTs' quotes are informative about the value of the security, consistent with our evidence. At this point, it is useful to mention that because the

4.4.4 *The information content of quotes on days of high/low volatility*

As explained previously, due to overnight and early morning news, the preopening auction often involves sharp price adjustments (e.g., [Moshirian et al., 2012](#)). To further investigate this feature, we replicate our analysis over days of high/low price adjustments relative to the previous close. For each stock, we split the sample period as follows. Using the full book prices, we consider a) 124 days (i.e., half of the sample days) that exhibit higher close-to-open logarithmic differences in absolute value (i.e., latent volatility), and b) 124 days of low close-to-open logarithmic differences in absolute value. Then, we run two separate regressions, according to equation (16), one for the sample period of low volatility (hereafter, denoted by *LOW*), and one for the sample period of high volatility (hereafter, denoted by *HIGH*). Figure 8 summarizes our results. In the top graph, we plot the across stocks average b coefficient, obtained from equation (16), for HFTs and non-HFTs, for the two periods. In the middle graph, we plot the corresponding difference in the inverse of the RMSE, ΔHN , which is used to assess the presence of informed HFTs, according to proposition (3.2). Additionally, in the bottom graph, we illustrate the resulting t -statistic from the test that the across stocks average ΔHN is statistically equal to zero. Taking these results into account, we conclude that HFTs are present in the market minutes before the opening (i.e., the corresponding t -statistics are greater than 1.96 in absolute value), on days of high volatility (i.e., in the *HIGH* period). By contrast, on days of low volatility (i.e., in the *LOW* period), the average ΔHN is not statistically significant and, therefore, we cannot derive an assertive conclusion that informed HFTs are present in the market.

[Figure 8 about here]

4.4.5 *A robustness test for price reversals after the opening*

We have illustrated that HFTs' bidding activity enhances price efficiency prior to the opening of the market. In this respect, our results suggest that, on average, the opening prices accurately predict future liquidation values of stocks. To further validate this result, in this section we examine whether prices reverse immediately after the opening in the continuous session. If this is the case,

BEDOFIH data set does not include information about investors' IDs, it is not possible to track investors' transactions in the main session and, in turn, to accurately calculate their realized profits. Herein, following [Biais et al. \(1999\)](#) and [Madhavan and Panchapagesan \(2000\)](#), we argue that the stock value is better represented by the closing price (or the mid-day price), assuming that most of the daily new information has been incorporated into prices. Our approach is in line with the well-documented U-shaped intraday pattern of volatility and illiquidity in financial markets.

then it is possible that the preopening price changes are transitory. By contrast, an absence of price corrections, after the opening, strongly suggests that the preopening price adjustments are indeed permanent, reflecting changes in the underlying true value (efficient price) of the stock (Pagano et al., 2013).

To test for price reversals after the opening, for each stock we fix the intraday interval and regress the main session intraday returns on the preopening intraday returns (across days) as follows:

$$R_{09:00 \text{ to } 09:00+k}^i = \theta_0 + \theta_1 R_{08:30 \text{ to } 09:00}^i + \zeta, \quad (17)$$

where $R_{08:30 \text{ to } 09:00}^i$ is the preopening logarithmic return from the actual indicative price at 08:30 (when HFTs rapidly enter the market) to the opening price for stock i , at 09:00, and $R_{09:00 \text{ to } 09:00+k}^i$ is the main session logarithmic return from the 09:00 opening price to the 09:00+ k main session price, with $k = 5, 10, 15, \dots, 60$ minutes after the opening. As a proxy of the main session price, we use the mid-point of the order book, $(A_1 + B_1)/2$, where A_1 and B_1 are the best sell and buy limits, respectively (Hasbrouck, 1991a). If there is a price correction after the opening, then θ_1 should be statistically significant and negative.

[Table 5 about here]

Results for the estimated θ_1 coefficients are presented in Table 5. The majority of estimated θ_1 coefficients (out of the 36 stock-regressions) are positive and statistically significant at the 5% probability level, when using the OLS standard errors (Panel A). On the other hand, when using Newey–West corrected errors (Panel B), the majority of these estimated θ_1 coefficients are not significantly different from zero at the 5% level (standard errors increase after the correction). It is important, however, to notice that a considerable percentage of θ_1 coefficients are positive and significant for several minutes after the opening (for instance, 30.56% in the first five minutes), although this feature fades out as the continuous process evolves.

Based on these results, we infer that, on average, there is little evidence to support a price reversal mechanism after the opening auction. Therefore, HFTs' quotes play an important role in price efficiency at the opening.

5 Conclusions

This paper contributes to the current debate in the microstructure literature about the effect of high-frequency trading on price quality by addressing an important question: Are HFTs informed about assets' true values? Existing empirical findings suggest that HFTs' quotes are informative in the very short-term (e.g., for a few seconds). Inferring truly informative price shocks within such short trading intervals, however, is questionable. We shed further light on this issue by examining the call auction environment, in which traders are allowed to gather and analyze public information for an extended intraday period of time, during which there is no actual trading.

We propose an empirical test to detect informed HFTs in the call market. To develop this test, we begin with the [Kyle \(1989\)](#) noisy rational expectations framework, for call auctions, which incorporates strategic trading. We extend this model to include high-frequency trading and, subsequently, to derive testable conditions under which informed HFTs can be seen to be present in the call market. Our test is based on a comparison of the information contained in the supply/demand schedules submitted by HFTs, non-HFTs, and the entire trading public.

To conduct our empirical test, we use a set of data from the Paris stock market's opening call. Our empirical results indicate that early preopening orders submitted by HFTs are rather noisy. As the opening time approaches, however, HFTs significantly increase their bidding intensity in terms of: i) the number of submitted orders, ii) order size, and iii) price aggressiveness. Our test results indicate that a few minutes before the opening informed HFTs are present in the market, improving price efficiency with their bids. Our evidence from the Paris market is consistent with recent empirical studies of the trading process in U.S. securities (NASDAQ and NYSE), which find that HFTs' quotes contain important information about asset values ([Carrion, 2013](#); [Brogaard et al., 2014](#)).

Our analysis adds to a growing strand in the microstructure literature that examines the characteristics of high-frequency trading in electronic call markets and the effect that HFTs have on the quality of clearing prices. The call auction mechanism is a useful tool for revealing prices after extended periods of no trade (e.g., at the opening of stock markets) or after periods of market instability (e.g, trading halts). Opening and closing prices are often used by portfolio managers for benchmark trading, derivative settlement activities, and portfolio evaluation. Therefore, an understanding of whether high-frequency trading algorithms efficiently process pre-trade information, contributing to

the price discovery process, is of critical importance. Our paper provides direct evidence that HFTs are important market participants in the Paris opening call, holding accurate information about asset values.

References

- Admati, A.R., 1985. A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica: Journal of the Econometric Society* 53, 629–657.
- Anagnostidis, P., Kanas, A., Papachristou, G., 2015. Information revelation in the Greek exchange opening call: Daily and intraday evidence. *Journal of International Financial Markets, Institutions and Money* 38, 167–184.
- Anand, A., Subrahmanyam, A., 2008. Information and the intermediary: Are market intermediaries informed traders in electronic markets? *Journal of Financial and Quantitative Analysis* 43, 1–28.
- Barclay, M.J., Hendershott, T., 2003. Price discovery and trading after hours. *Review of Financial Studies* 16, 1041–1073.
- Barclay, M.J., Hendershott, T., 2008. A comparison of trading and non-trading mechanisms for price discovery. *Journal of Empirical Finance* 15, 839–849.
- Barclay, M.J., Warner, J.B., 1993. Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics* 34, 281–305.
- Baruch, S., Panayides, M., Venkataraman, K., 2017. Informed trading and price discovery before corporate events. *Journal of Financial Economics* 125, 561–588.
- Bellia, M., Pelizzon, L., Subrahmanyam, M.G., Uno, J., Yuferova, D., 2017. Coming Early to the Party. SAFE Working Paper No. 182, 20 September. Available at SSRN 3038699.
- Bellia, M., Pelizzon, L., Subrahmanyam, M.G., Uno, J., Yuferova, D., 2020. Low-Latency Trading and Price Discovery: Evidence from the Tokyo Stock Exchange in the Pre-Opening and Opening Periods. SAFE Working Paper No. 144, 22 June. Available at SSRN 3619823.
- Biais, B., Foucault, T., 2014. HFT and market quality. *Bankers, Markets & Investors* 128, 5–19.
- Biais, B., Foucault, T., Moinas, S., 2015. Equilibrium fast trading. *Journal of Financial Economics* 116, 292–313.
- Biais, B., Hillion, P., Spatt, C., 1995. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *The Journal of Finance* 50, 1655–1689.
- Biais, B., Hillion, P., Spatt, C., 1999. Price discovery and learning during the preopening period in the Paris Bourse. *Journal of Political Economy* 107, 1218–1248.
- Boehmer, E., Fong, K.Y., Wu, J.J., 2020. Algorithmic Trading and Market Quality: International Evidence. Working Paper, 15 May. Available at SSRN 2022034.
- Boulatov, A., Hendershott, T., Livdan, D., 2012. Informed trading and portfolio returns. *Review of Economic Studies* 80, 35–72.
- Boussetta, S., Lescourret, L., Moinas, S., 2020. Does a pre-open matter in fragmented markets? Working Paper, 12 May. Available at SSRN 2939502.
- Bouveret, A., Guillaumie, C., Roqueiro, C.A., Winkler, C., Nauhaus, S., 2014. High-frequency

- trading activity in EU equity markets. ESMA Report on Trends, Risks and Vulnerabilities 1, 41–47.
- Brogaard, J., Hendershott, T., Riordan, R., 2014. High-frequency trading and price discovery. *The Review of Financial Studies* 27, 2267–2306.
- Brogaard, J., Hendershott, T., Riordan, R., 2019. Price discovery without trading: Evidence from limit orders. *The Journal of Finance* 74, 1621–1658.
- Cao, C., Ghysels, E., Hatheway, F., 2000. Price discovery without trading: Evidence from the Nasdaq preopening. *The Journal of Finance* 55, 1339–1365.
- Carrion, A., 2013. Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets* 16, 680–711.
- Cespa, G., Foucault, T., 2013. Sale of price information by exchanges: does it promote price discovery? *Management Science* 60, 148–165.
- Chakravarty, S., 2001. Stealth-trading: Which traders' trades move stock prices? *Journal of Financial Economics* 61, 289–307.
- Chlistalla, M., 2012. High-frequency trading: better than its reputation? Research briefing, Deutsche Bank Research.
- Ciccotello, C.S., Hatheway, F.M., 2000. Indicating ahead: Best execution and the Nasdaq preopening. *Journal of Financial Intermediation* 9, 184–212.
- Collin-Dufresne, P., Fos, V., 2015. Do prices reveal the presence of informed trading? *The Journal of Finance* 70, 1555–1582.
- Comerton-Forde, C., Rydge, J., 2006. The influence of call auction algorithm rules on market efficiency. *Journal of Financial Markets* 9, 199–222.
- De Jong, F., Rindi, B., 2009. *The microstructure of financial markets*. Cambridge University Press.
- Easley, D., Kiefer, N.M., O'hara, M., Paperman, J.B., 1996. Liquidity, information, and infrequently traded stocks. *The Journal of Finance* 51, 1405–1436.
- Foucault, T., Hombert, J., Roşu, I., 2016. News trading and speed. *The Journal of Finance* 71, 335–382.
- Fox, M.B., Glosten, L.R., Rauterberg, G.V., 2018. Informed trading and its regulation. *Journal of Corporation Law* 43, 817–860.
- Gao, L., Han, Y., Li, S.Z., Zhou, G., 2018. Market intraday momentum. *Journal of Financial Economics* 129, 394–414.
- Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.
- Grossman, S.J., Stiglitz, J.E., 1976. Information and competitive price systems. *The American Economic Review* 66, 246–253.

- Grossman, S.J., Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. *The American Economic Review* 70, 393–408.
- Hasbrouck, J., 1991a. Measuring the information content of stock trades. *The Journal of Finance* 46, 179–207.
- Hasbrouck, J., 1991b. The summary informativeness of stock trades: An econometric analysis. *The Review of Financial Studies* 4, 571–595.
- Hasbrouck, J., Saar, G., 2013. Low-latency trading. *Journal of Financial Markets* 16, 646–679.
- Hauser, S., Kamara, A., Shurki, I., 2012. The effects of randomizing the opening time on the performance of a stock market under stress. *Journal of Financial Markets* 15, 392–415.
- Hellwig, M.F., 1980. On the aggregation of information in competitive markets. *Journal of Economic Theory* 22, 477–498.
- Hendershott, T., Livdan, D., Schürhoff, N., 2015. Are institutions informed about news? *Journal of Financial Economics* 117, 249–287.
- Hillion, P., Suominen, M., 2004. The manipulation of closing prices. *Journal of Financial Markets* 7, 351–375.
- Hirschey, N., 2019. Do high-frequency traders anticipate buying and selling pressure? Working paper, 8 October. Available at SSRN 2238516.
- Huang, R.D., Stoll, H.R., 1997. The components of the bid-ask spread: A general approach. *The Review of Financial Studies* 10, 995–1034.
- Jarrow, R.A., Protter, P., 2012. A dysfunctional role of high frequency trading in electronic markets. *International Journal of Theoretical and Applied Finance* 15, 1–15.
- Kalay, A., Sade, O., Wohl, A., 2004. Measuring stock illiquidity: An investigation of the demand and supply schedules at the TASE. *Journal of Financial Economics* 74, 461–486.
- Kalay, A., Wohl, A., 2009. Detecting liquidity traders. *Journal of Financial and Quantitative Analysis* 44, 29–54.
- Kovalenkov, A., Vives, X., 2014. Competitive rational expectations equilibria without apology. *Journal of Economic Theory* 149, 211–235.
- Kyle, A.S., 1989. Informed speculation with imperfect competition. *The Review of Economic Studies* 56, 317–355.
- Madhavan, A., Panchapagesan, V., 2000. Price discovery in auction markets: A look inside the black box. *Review of Financial Studies* 13, 627–658.
- Medrano, L.A., Vives, X., 2001. Strategic behavior and price discovery. *RAND Journal of Economics* 32, 221–248.
- Megarbane, N., Saliba, P., Lehalle, C.A., Rosenbaum, M., 2017. The behavior of high-frequency traders under different market stress scenarios. *Market Microstructure and Liquidity* 3, 1850005.

- Menkhoff, L., Schmeling, M., 2010. Whose trades convey information? Evidence from a cross-section of traders. *Journal of Financial Markets* 13, 101–128.
- Moshirian, F., Nguyen, H.G.L., Pham, P.K., 2012. Overnight public information, order placement, and price discovery during the pre-opening period. *Journal of Banking & Finance* 36, 2837–2851.
- O’Hara, M., 2015. High frequency market microstructure. *Journal of Financial Economics* 116, 257–270.
- Pagano, M.S., Peng, L., Schwartz, R.A., 2013. A call auction’s impact on price formation and order routing: Evidence from the NASDAQ stock market. *Journal of Financial Markets* 16, 331–361.
- Pagano, M.S., Schwartz, R.A., 2003. A closing call’s impact on market quality at Euronext Paris. *Journal of Financial Economics* 68, 439–484.
- SEC, 2010. Concept Release on Equity Market Structure, Release No. 34-61358; File No. Technical Report. S7-02-10.
- Stoll, H.R., Whaley, R.E., 1990. Stock market structure and volatility. *The Review of Financial Studies* 3, 37–71.
- Van Kervel, V., 2015. Competition for order flow with fast and slow traders. *The Review of Financial Studies* 28, 2094–2127.

Tables

Table 1: Comparative statics analysis results.

Parameter	$Var(u p)^{-1}$
$Z \uparrow$	\downarrow
$M \uparrow$	\uparrow
$N \uparrow$	\uparrow
$\sigma_e \uparrow$	\downarrow
$\rho \uparrow$	\downarrow

Note: We report comparative statics results for the behavior of the efficiency of p with respect to the following model parameters: N (number of informed traders), M (number of uninformed traders), Z (number of noise traders), σ_e (standard deviation of the error of the private signal), and ρ (risk aversion parameter). We remark that the number of uninformed traders, L , implicitly affects price efficiency through the modification of price elasticity, λ_I .

Table 2: Preopening order flow for the CAC 40 sample.

Time	07:15–07:30	07:30–07:45	07:45–08:00	08:00–08:15	08:15–08:30	08:30–08:45	08:45–Open
OHFT							
Submit total	385.0	149.0	14,937.0	3,620.0	2,265.0	2,018,873.0	1,029,859.0
Submit average	0.0	0.0	1.6	0.4	0.2	220.8	112.6
Average size	1,807.0	813.3	2,173.6	673.9	823.3	192.1	667.9
Modify total	0.0	0.0	0.0	0.0	5.0	570.0	334,694.0
Modify average	0.0	0.0	0.0	0.0	0.0	0.1	36.6
Cancel total	6.0	10.0	6.0	145.0	1,166.0	3,620.0	279,430.0
Cancel average	0.0	0.0	0.0	0.0	0.1	0.4	30.6
MIXED							
Submit total	80.0	336.0	3,052.0	245,510.0	29,949.0	102,097.0	1,188,136.0
Submit average	0.0	0.0	0.3	26.8	3.3	11.2	129.9
Average size	2,776.0	3,729.1	1,486.7	2,095.0	2,432.8	1,824.3	5,938.6
Modify total	4.0	87.0	1,877.0	2,178.0	1,759.0	6,013.0	538,900.0
Modify average	0.0	0.0	0.2	0.2	0.2	0.7	58.9
Cancel total	4.0	9.0	812.0	24,847.0	14,088.0	9,346.0	304,282.0
Cancel average	0.0	0.0	0.1	2.7	1.5	1.0	33.3
non-HFT							
Submit total	546,131.0	475,968.0	198,644.0	147,416.0	77,681.0	958,569.0	295,765.0
Submit average	59.7	52.1	21.7	16.1	8.5	104.8	32.3
Average size	216.8	187.0	338.9	288.8	653.2	779.8	1,373.8
Modify total	587.0	599.0	280.0	374.0	278.0	2,299.0	160,165.0
Modify average	0.1	0.1	0.0	0.0	0.0	0.3	17.5
Cancel total	54,282.0	83,630.0	83,286.0	20,355.0	18,608.0	18,584.0	54,548.0
Cancel average	5.9	9.1	9.1	2.2	2.0	2.0	6.0

Note: We report the average and total (sum), across days and stocks, number of new submissions, revisions and cancellations during the preopening and in 15 minute intervals, for each type of trader in the data set; OHFTs, MIXED traders, and non-HFTs. Zero averages (0.0) represent very small values (< 0.1). For submissions, the, across days and stocks, average order size is also reported.

Table 3: Opening statistics for the 36 stock sample from the CAC 40 Index.

Security	BUY OHFT (%)	BUY MIXED (%)	BUY non-HFT (%)	SELL OHFT (%)	SELL MIXED (%)	SELL non-HFT (%)	Trades (%)	Trades average	Volume average	Value average
Accor	4.3	60.8	35.0	4.4	48.2	47.4	1.2	46.9	8,953.6	262,413.0
Air Liquide	5.9	39.1	55.1	3.1	34.7	62.2	1.9	124.1	13,806.2	1,343,993.7
Axa	4.4	45.3	50.3	4.0	36.0	60.0	1.5	145.5	96,130.2	1,533,514.0
BNP Paribas	4.0	42.4	53.6	2.9	34.8	62.3	1.1	179.3	57,957.6	2,728,559.1
Cap Gemini	6.9	46.0	47.1	6.2	40.9	52.9	1.2	58.5	8,940.1	356,822.5
Carrefour	4.6	49.8	45.6	5.2	45.8	49.0	1.3	89.4	30,825.1	709,855.6
Credit Agricole	6.8	20.3	72.9	5.5	20.7	73.8	2.1	190.0	126,248.8	959,861.9
Danone	3.0	66.9	30.0	3.0	50.7	46.4	1.3	92.5	27,963.2	1,529,767.4
Lvmh	4.5	49.2	46.3	4.4	48.4	47.2	1.6	106.8	13,617.0	1,838,125.0
Michelin Nom.	6.8	49.7	43.6	3.1	31.8	65.1	1.6	103.9	10,707.6	775,199.2
L'Oreal	5.4	54.4	40.2	4.5	45.8	49.8	1.5	80.6	9,845.7	1,209,364.8
Pernod Ricard	7.2	56.7	36.1	4.0	52.6	43.4	1.2	61.9	7,246.9	654,519.6
Kering	4.9	54.1	41.0	4.9	45.9	49.2	1.4	47.8	2,992.8	494,872.3
Saint Gobain	6.2	54.7	39.1	4.5	43.4	52.2	1.1	79.3	22,456.8	748,897.3
Schneider Electric	5.0	57.0	38.0	5.1	53.3	41.7	0.9	78.9	19,869.7	1,176,372.9
Total	3.2	44.3	52.5	3.7	36.0	60.3	1.4	189.6	82,860.6	3,339,932.5
Veolia Environn.	6.1	37.8	56.1	6.3	37.0	56.7	1.4	79.4	34,383.7	374,904.4
Vinci	3.6	45.5	50.9	4.0	34.3	61.8	1.6	115.9	25,525.7	1,016,483.0
Vivendi	6.4	45.6	47.9	5.1	41.2	53.7	1.8	120.8	73,438.7	1,211,928.8
Essilor Intl	5.1	49.0	45.9	3.8	52.7	43.5	1.4	75.7	8,571.6	689,196.2
SAFRAN	3.2	40.5	56.3	2.6	41.2	56.2	1.8	78.8	11,744.4	485,001.1
Bouygues	6.7	44.7	48.6	6.1	43.6	50.3	1.2	54.8	11,908.5	278,165.5
Publicis Groupe	5.6	58.7	35.7	4.9	53.8	41.3	0.9	44.0	6,749.9	389,673.0
Societe Generale	3.8	35.6	60.5	3.9	30.7	65.4	1.0	170.5	63,474.1	2,063,893.2
Technip	7.6	40.7	51.7	4.9	48.7	46.4	1.5	82.4	8,392.0	664,720.3
Renault Group	4.6	51.8	43.7	5.8	42.0	52.2	1.0	77.9	15,921.1	888,674.5
Airbus Group	2.6	41.0	56.4	1.9	24.1	74.0	1.8	194.4	40,984.7	1,772,246.1
Lafarge	6.2	38.9	54.8	6.0	33.3	60.7	1.5	84.8	8,687.1	433,775.4
Stmicroelectronics	9.5	41.7	48.8	8.5	46.9	44.7	1.0	41.9	32,015.2	203,453.4
EDF	3.0	40.2	56.8	3.4	21.8	74.8	2.4	117.4	24,131.0	465,484.1
Alstom	5.6	31.8	62.6	6.4	37.2	56.4	1.7	103.0	21,355.4	599,144.3
Vallourec	7.2	38.6	54.2	6.3	38.7	55.0	1.4	67.9	8,679.3	365,233.8
Engie	3.9	46.9	49.3	2.9	30.2	66.9	2.4	169.8	71,119.6	1,164,081.3
Legrand	3.6	61.7	34.7	2.7	53.1	44.1	1.3	41.3	7,831.7	290,675.1
Orange	4.3	35.6	60.1	3.4	32.5	64.1	1.6	168.0	127,862.4	1,069,817.0
Sanofi	4.4	50.5	45.1	3.5	48.0	48.5	1.1	147.2	52,700.8	4,042,025.7
Average across stocks	5.2	46.3	48.5	4.5	40.5	55.0	1.4	103.1	33,219.4	1,059,184.6
Out of sample	Platform negotiated									
Solvay	Belgium									
Gemalto	the Netherlands									
Unibail-Rodamco	the Netherlands									
ArcelorMittal Reg	Luxembourg									

Note: The following statistics are reported for the opening trade: i) OHFTs', MIXED traders', and non-HFTs' percentage of orders on the buy/sell side of the market (the percentage for each of the three categories and for each market side is calculated as the number of flagged (e.g., OHFT) messages relative to the total number of opening trade messages), ii) percentage of opening trades relative to total daily trades, iii) daily average number of trades, iv) daily average clearing volume (shares), and v) daily average trading value (Euros). Four stocks from the CAC 40 Index are excluded from our analysis as they are not negotiated on the Paris platform.

Table 4: Comparison of price efficiency during the preopening and at the opening.

Time	ΔFH (%+)	$\overline{\Delta FH}$	$t_{\Delta FH}$	ΔFN (%+)	$\overline{\Delta FN}$	$t_{\Delta FN}$	ΔHN (%+)	$\overline{\Delta HN}$	$t_{\Delta HN}$
08:31	58.33	0.14	1.01	61.11	0.01	0.22	50.00	-0.13	-0.83
08:32	61.11	0.17	1.13	61.11	0.06	0.84	55.56	-0.11	-0.67
08:33	61.11	0.16	1.00	63.89	0.08	1.00	50.00	-0.08	-0.46
08:34	63.89	0.20	1.33	63.89	0.08	1.00	50.00	-0.12	-0.75
08:35	63.89	0.20	1.27	61.11	0.09	1.10	55.56	-0.11	-0.64
08:36	61.11	0.27	1.80	63.89	0.11	1.43	50.00	-0.15	-0.99
08:37	58.33	0.17	1.25	66.67	0.15*	2.15	52.78	-0.02	-0.11
08:38	55.56	0.10	0.66	69.44	0.27*	3.63	58.33	0.17	1.06
08:39	55.56	0.10	0.65	69.44	0.32*	3.87	61.11	0.22	1.36
08:40	58.33	0.13	0.89	72.22	0.35*	4.10	58.33	0.22	1.38
08:41	58.33	0.12	0.78	77.78	0.40*	4.44	58.33	0.28	1.66
08:42	61.11	0.14	0.89	77.78	0.44*	5.15	61.11	0.30	1.71
08:43	52.78	0.19	1.12	88.89	0.51*	5.99	63.89	0.33	1.79
08:44	58.33	0.31	1.85	86.11	0.62*	5.99	63.89	0.31	1.70
08:45	61.11	0.28	1.70	86.11	0.63*	5.67	63.89	0.34	1.77
08:46	63.89	0.38*	2.23	83.33	0.70*	5.81	63.89	0.32	1.80
08:47	61.11	0.38*	2.15	86.11	0.72*	6.16	61.11	0.34	1.75
08:48	61.11	0.37*	2.19	86.11	0.74*	6.26	61.11	0.37	1.86
08:49	61.11	0.33	1.88	83.33	0.81*	6.33	61.11	0.48*	2.05
08:50	61.11	0.25	1.21	86.11	0.85*	6.12	58.33	0.60*	2.18
08:51	86.11	0.82*	6.03	86.11	1.04*	6.21	63.89	0.22	1.25
08:52	88.89	0.96*	5.96	91.67	1.25*	6.88	63.89	0.29	1.47
08:53	94.44	1.10*	6.30	91.67	1.24*	7.17	61.11	0.14	0.74
08:54	83.33	1.18*	5.47	91.67	1.46*	7.19	58.33	0.27	1.38
08:55	88.89	1.53*	6.53	91.67	1.64*	6.61	52.78	0.11	0.51
08:56	88.89	1.65*	6.21	91.67	1.78*	7.07	44.44	0.13	0.51
08:57	91.67	1.66*	6.58	97.22	2.17*	8.84	58.33	0.52*	2.00
08:58	86.11	0.96*	5.79	97.22	2.84*	10.29	86.11	1.88*	6.32
08:59	75.00	0.62*	3.77	100.00	3.96*	10.97	91.67	3.34*	7.77
09:00	88.89	1.56*	6.37	100.00	6.61*	11.11	97.22	5.05*	8.82

Note: Comparison of the informational content of $p_{FULLBOOK}$, p_{HFT} , and $p_{non-HFT}$ from 8:30 to 9:00 (opening) in 1-minute frequency. We report statistics for the following differences: i) $\Delta FH = RMSE_{FULLBOOK}^{-1} - RMSE_{HFT}^{-1}$, ii) $\Delta FN = RMSE_{FULLBOOK}^{-1} - RMSE_{non-HFT}^{-1}$ and iii) $\Delta HN = RMSE_{HFT}^{-1} - RMSE_{non-HFT}^{-1}$. For each case, reported statistics include the across stocks percentage of positive differences, the across stocks average difference, and the t-statistic obtained from a t-test for the null hypothesis that the across stocks average difference is not different from zero. Asterisks denote significance at the 5% probability level. Inverse $RMSE$ s are obtained from the estimation of the unbiasedness regression $R_{cc} = a + bR_{ct} + \epsilon$ (equation (16)) using the three prices, $p_{FULLBOOK}$, p_{HFT} , and $p_{non-HFT}$ in the calculation of R_{ct} .

Table 5: Price reversals after the opening.

Panel A					
k (minutes)	Average θ_1	Positive and significant θ_1 (% out 36 stocks)	Negative and significant θ_1 (% out 36 stocks)	Not significant θ_1 (% out 36 stocks)	
5	0.0334	91.67	0.00	8.33	
10	0.0345	86.11	0.00	13.89	
15	0.0315	88.89	0.00	11.11	
20	0.0288	88.89	0.00	11.11	
25	0.0322	88.89	0.00	11.11	
30	0.0256	94.44	0.00	5.56	
35	0.0324	91.67	0.00	8.33	
40	0.0329	91.67	0.00	8.33	
45	0.0364	97.22	0.00	2.78	
50	0.0330	94.44	0.00	5.56	
55	0.0326	94.44	0.00	5.56	
60	0.0373	88.89	0.00	11.11	
Panel B					
k (minutes)	Average θ_1	Positive and significant θ_1 (% out 36 stocks)	Negative and significant θ_1 (% out 36 stocks)	Not significant θ_1 (% out 36 stocks)	
5	0.0334	30.56	2.78	66.67	
10	0.0345	30.56	5.56	63.89	
15	0.0315	25.00	2.78	72.22	
20	0.0288	16.67	5.56	77.78	
25	0.0322	25.00	2.78	72.22	
30	0.0256	8.33	2.78	88.89	
35	0.0324	13.89	2.78	83.33	
40	0.0329	8.33	2.78	88.89	
45	0.0364	11.11	2.78	86.11	
50	0.0330	11.11	2.78	86.11	
55	0.0326	11.11	2.78	86.11	
60	0.0373	11.11	2.78	86.11	

Note: We fix the intraday interval and regress, across days, main session intraday returns ($R_{09:00\ to\ 09:00+k}^i$) on the preopening intraday returns ($R_{08:30\ to\ 09:00}^i$) in equation (17): $R_{09:00\ to\ 09:00+k}^i = \theta_0 + \theta_1 R_{08:30\ to\ 09:00}^i + \zeta$. We report the across stocks average θ_1 regression coefficient estimate, the percentage of statistically significant (positive or negative) θ_1 coefficient estimates at the 5% probability level, out of the 36 sample stock regressions, and the percentage (out of the 36 sample stock regressions) of coefficient estimates that are not significant (at the 5% probability level). Regressions are conducted over a range of post-opening periods: $k = 5, 10, 15, \dots, 60$ minutes after the opening. In Panel A statistical significance is assessed using the regression standard errors, while in Panel B significance is inferred using Newey–West errors corrected for serial correlation and heteroscedasticity.

Figures

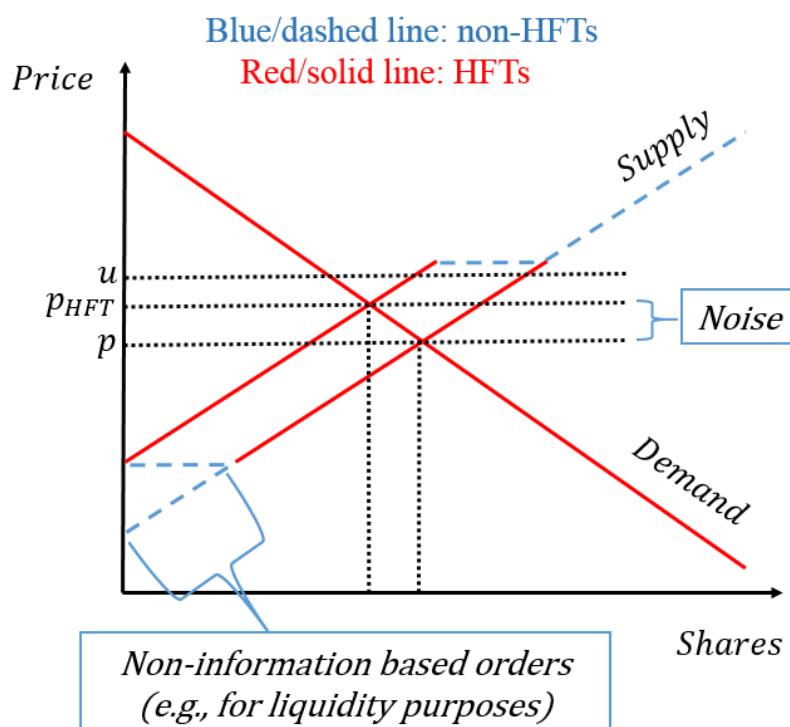


Figure 1: Calculation of p_{HFT} in a hypothetical supply–demand schedule submitted by non–HFTs (blue/dashed line) and HFTs (red/solid line). p is the equilibrium price calculated at the point where (all) traders’ aggregate net demand is zero. p_{HFT} is calculated at the point where the aggregate net demand submitted by HFTs is zero. u is the future liquidation value of the stock.

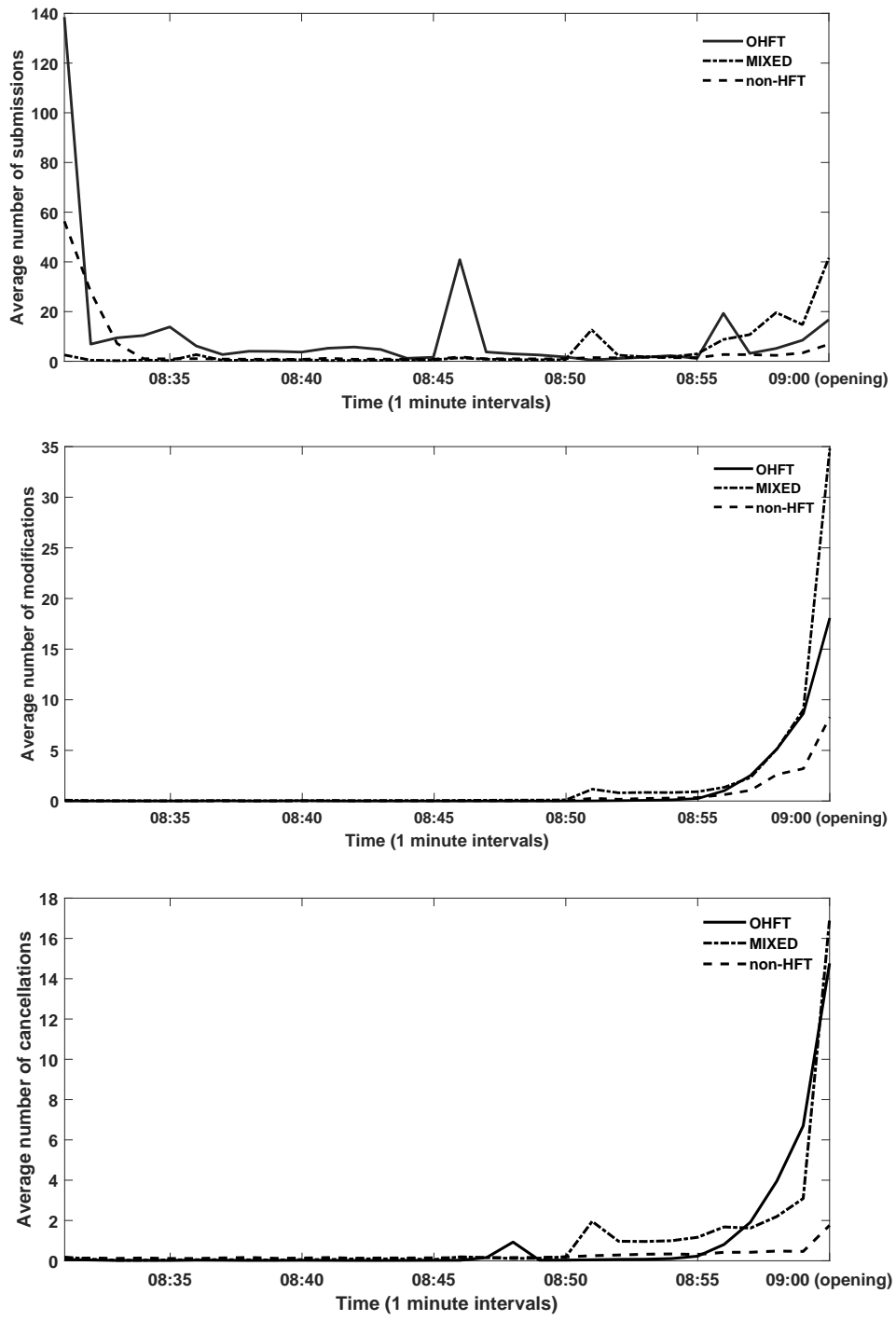


Figure 3: The across stocks and days average number of submissions, modifications and cancellations during the last 30 minutes of the preopening, in 1-minute intervals.

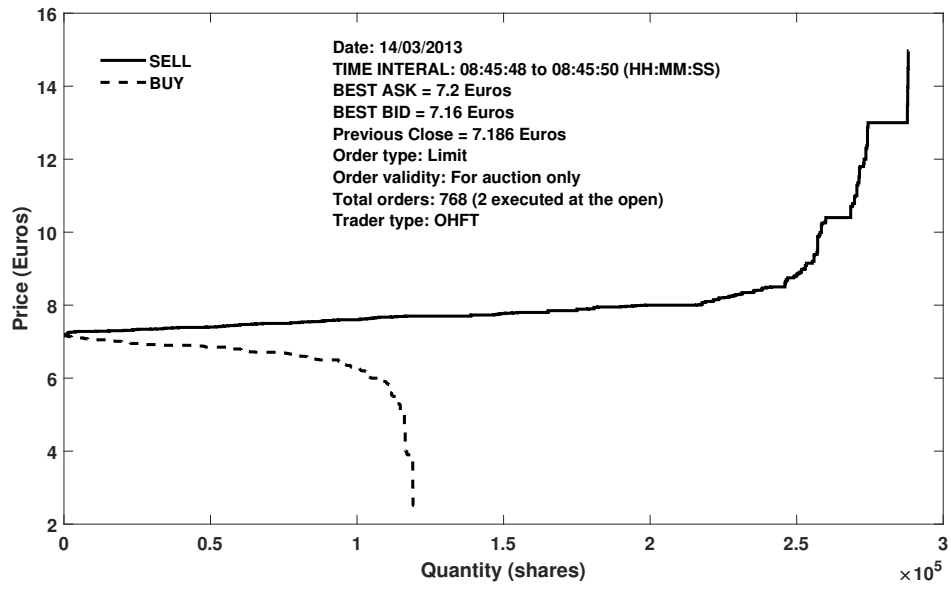


Figure 4: A sequence of orders submitted at 08:45 for Credit Agricole stock on 14/03/2013.

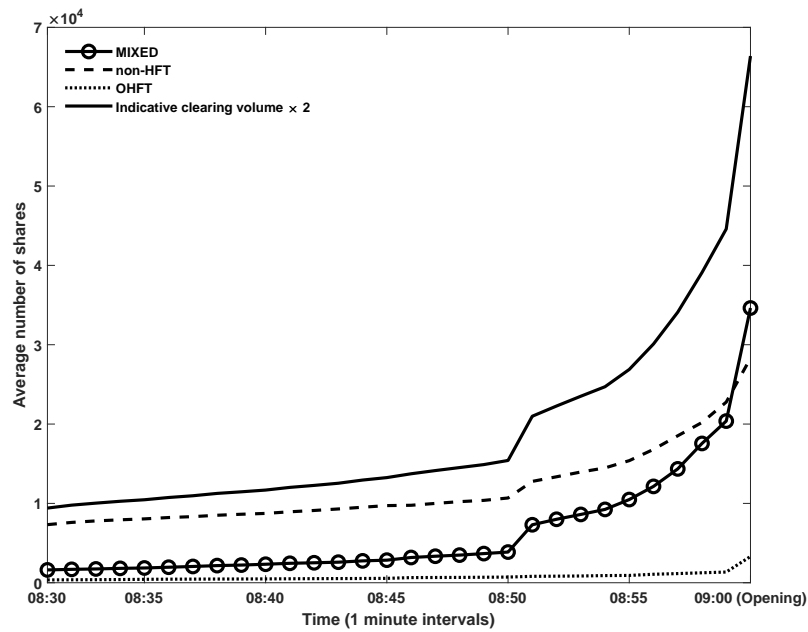
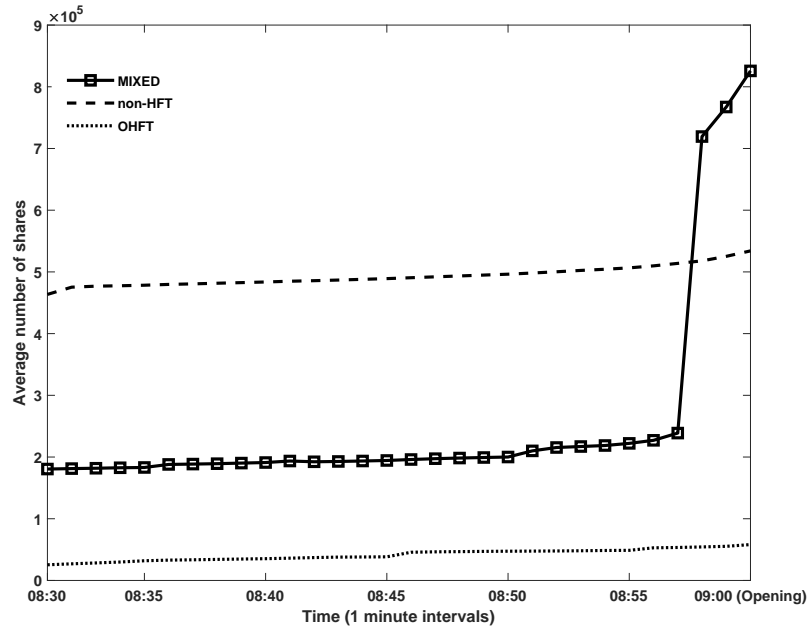


Figure 5: Top graph: The across stocks and days average number of available shares (buy and sell) on the order book due to OHFTs', MIXED traders', and non-HFTs' quotes. Bottom graph: The across stocks and days average number of shares on the executable side of the order book (buy and sell) due to OHFTs', MIXED traders', and non-HFTs' quotes. In addition, we plot the indicative clearing volume multiplied by two. Given that the indicative/clearing price is p , executable buy (sell) orders are those with a limit price higher than or equal to (lower than or equal to) p . Market and at-the-open orders which have no price preference are considered as infinitely aggressive and thus they have the highest execution priority.

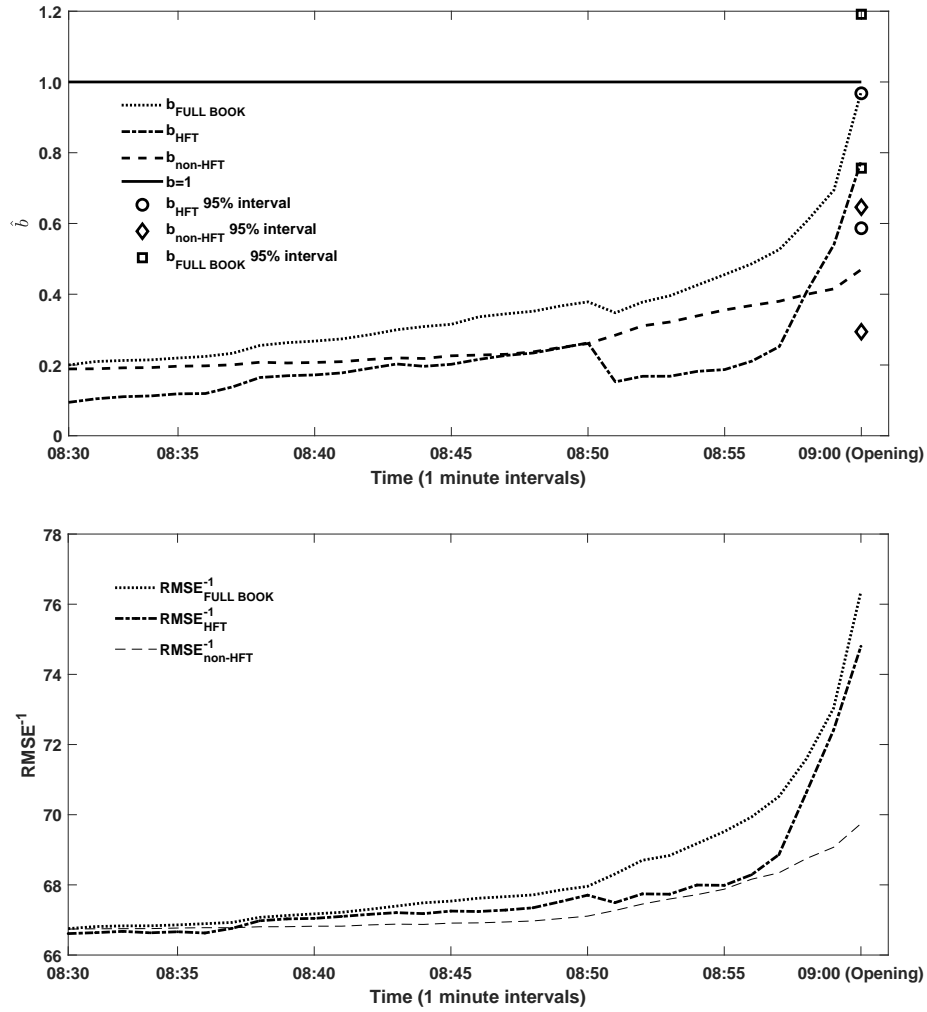


Figure 6: Upper graph: The across stocks average estimated unbiasedness regression b coefficient during the preopening, obtained using the three prices: p_{HFT} , $p_{non-HFT}$ and $p_{FULLBOOK}$ in the calculation of R_{ct} in equation (16). Bottom graph: The corresponding average of the inverse of the $RMSE$.

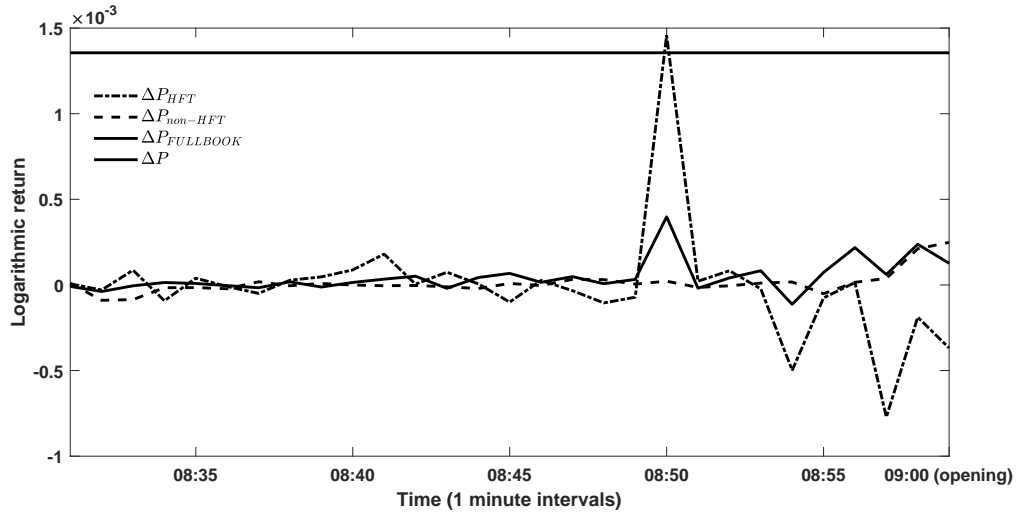


Figure 7: The across days and stocks average 1-minute logarithmic price change, Δp_k , $k = 1, \dots, 29$, during the last 30 minutes of the preopening, calculated using the three prices: p_{HFT} , $p_{non-HFT}$, and $p_{FULLBOOK}$. We also plot the across days and stocks average total price change from 08:30 to 09:00, $\Delta p^{FULLBOOK}$ (i.e., the horizontal line), calculated using the full book price, $p_{FULLBOOK}$.

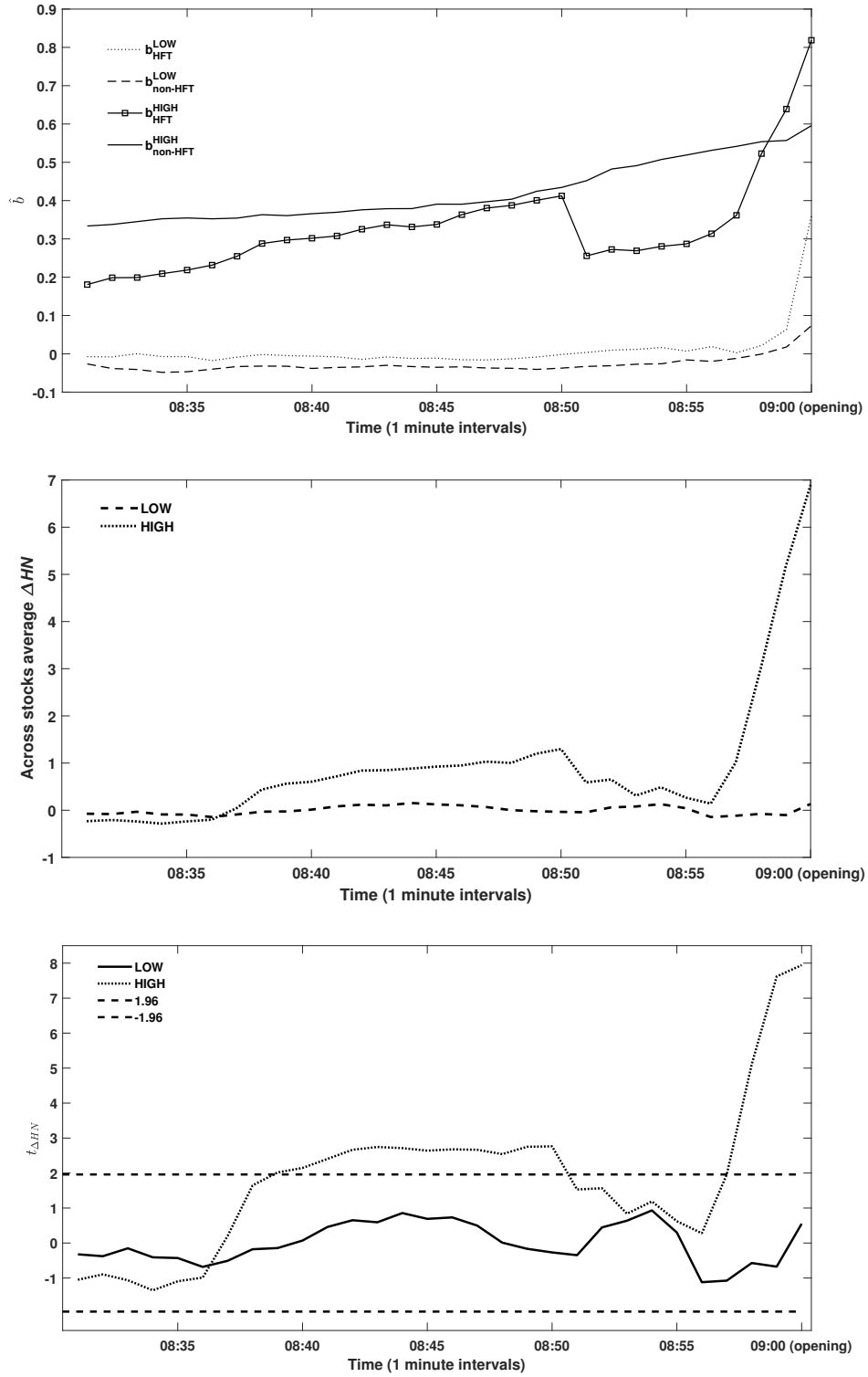


Figure 8: Top graph: The across stocks average b coefficient, obtained from equation (16), for HFTs and non-HFTs, for the two volatility periods (HIGH/LOW). Middle graph: The corresponding difference in the inverse of the RMSE, $\Delta HN = RMSE_{HFT}^{-1} - RMSE_{non-HFT}^{-1}$, which is used to assess the presence of informed HFTs, according to proposition (3.2). Bottom graph: The calculated t-statistic for the test that the across stocks average ΔHN is statistically equal to zero.

Appendix A Informed and uninformed traders' problem

To facilitate an understanding of our analytic framework, in this Appendix we solve the optimization problem for informed and uninformed traders, as presented in Kyle (1989). We derive formulas for traders' expectations about the fair value of the asset (u) and, subsequently, calculate the quantity submitted by each trader-type (i.e., X and Y).

Informed traders maximize their expected utility,

$$E[U(\pi_I)|(p, \iota)] = E[-e^{-\rho\pi_I}|(p, \iota)] \quad (\text{A.1})$$

where $\pi_I = X(u - p)$ is the terminal wealth, and p is the clearing price. Note that since u is normally distributed, π_I is also normal

$$\pi_I \sim N(X(E(u|p, \iota) - p), X^2 \text{Var}(u|p, \iota)). \quad (\text{A.2})$$

Using the properties of normal distribution, it is straightforward to show that maximizing the expected utility is equivalent to maximizing the following quantity (certainty equivalent)

$$E[\pi_I|(p, \iota)] - \frac{\rho}{2} \text{Var}[\pi_I|(p, \iota)]. \quad (\text{A.3})$$

Thus, to maximize expected utility, informed traders use the following first order condition with respect to X :

$$E(u|\iota, p) + \frac{\partial E(u|p, \iota)}{\partial X} - p - \frac{\partial p}{\partial X} X - \rho X \text{Var}(u|\iota, p) - \frac{\rho}{2} X^2 \frac{\partial \text{Var}(u|p, \iota)}{\partial X} = 0. \quad (\text{A.4})$$

Using the projection theorem for normal variables, informed traders can compute the conditional expectation and variance of u , with respect to p and ι , as follows:

$$E(u|\iota, p) = \iota \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = \delta \iota = E(u|\iota) \quad (\text{A.5})$$

and

$$\text{Var}(u|\iota, p) = \sigma_e^2 \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = \delta \sigma_e^2 = \text{Var}(u|\iota) \quad (\text{A.6})$$

where $\delta = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$. Notice that, since ι is common to all informed agents, it is a sufficient statistic for the market price; that is, informed agents cannot learn something more from the market price and, therefore, form their demand based solely on their exogenous private signal.³⁵

Equations (A.5) and (A.6) imply that $\frac{\partial E(u|p, \iota)}{\partial X} = 0$ and $\frac{\partial \text{Var}(u|p, \iota)}{\partial X} = 0$. Thus, the first order condition for informed traders becomes:

$$E(u|\iota, p) - p - \frac{\partial p}{\partial X} X - \rho X \text{Var}(u|\iota, p) = 0. \quad (\text{A.7})$$

By solving the aforementioned equation, the demand function for informed traders assumes the following form,

$$X = \frac{E(u|\iota, p) - p}{\lambda_I + \rho \text{Var}(u|\iota, p)}. \quad (\text{A.8})$$

Explicating the conditional moments, we obtain:

$$X = \frac{\delta}{\lambda_I + \rho \delta \sigma_e^2} \iota - \frac{1}{\lambda_I + \rho \delta \sigma_e^2} p, \quad (\text{A.9})$$

which, in view of our initial formulation for the demand schedules (at the end, expectations coincide with realizations), X , yields the following relations:

$$\gamma_I = \frac{1}{\lambda_I + \rho \delta \sigma_e^2}, \quad (\text{A.10})$$

$$\beta = \frac{\delta}{\lambda_I + \rho \delta \sigma_e^2}. \quad (\text{A.11})$$

Notice, also, that $\mu_I = 0$, such that $X = -\gamma_I p + \beta \iota + \mu_I$.

Uninformed traders do not hold private information and thus use their prior beliefs, conditional on the observed indicative price, p , to decide about their future investments. To extract information from p , each uninformed agent conjectures the following:

³⁵In Kyle (1989), each informed agent possesses a small piece of information and thus he/she submits his/her demand based both on both p and ι . We can extend our framework to include heterogeneous private information without qualitatively altering our analysis.

- Informed agents' demand is X , described by equations (A.9) and (A.10). In other words, uninformed agents are able to guess X , since informed agents' demand is based exclusively on their private signal and, therefore, informed agents do not reformulate their demand in the process of convergence to the equilibrium (De Jong and Rindi, 2009).
- All other uninformed agents submit a demand of the form Y , as described previously.

By assuming the above, and using the market clearing condition (1), uninformed agents can extract the following signal:

$$\Theta = \iota + \frac{1}{N\beta} ZQ = \delta^* p, \quad (\text{A.12})$$

where

$$\delta^* = \frac{N\gamma_I + M\gamma_U}{N\beta}. \quad (\text{A.13})$$

It is straightforward to show that $E(u|p) = E(u|\Theta)$ and $Var(u|p) = Var(u|\Theta)$. That is, p and Θ are informationally equivalent. Using Θ , uninformed traders can now compute their demand function, Y .

Similar to the informed traders' problem, uninformed traders aim to maximize their expected wealth, $E[\pi_U|\Theta]$, given the extracted signal Θ . This is equivalent to maximizing: $E[\pi_U|\Theta] - (\rho/2)(Var[\pi_U|\Theta])$, which leads to the following first order condition with respect to Y :

$$E(u|\Theta) - p - \frac{\partial p}{\partial Y} Y - \rho Y Var(u|\Theta) = 0. \quad (\text{A.14})$$

Notice here that, because Θ is a sufficient statistic for the market price, p (that is, $E(u|p) = E(u|\Theta)$ and $Var(u|p) = Var(u|\Theta)$), it follows immediately that $\partial E(u|\Theta)/\partial Y = 0$ and $\partial Var(u|\Theta)/\partial Y = 0$. Solving for the demand Y , submitted by uninformed traders, we obtain

$$Y = \frac{E(u|\Theta) - p}{\lambda_U + \rho Var(u|\Theta)}. \quad (\text{A.15})$$

Using the projection theorem to calculate $E(u|\Theta)$, it is straightforward to show that equation (A.15) can be rearranged as follows:

$$Y = -\frac{1 - D\delta^*}{\lambda_U + \rho Var(u|\Theta)} p \quad (\text{A.16})$$

with

$$D = \frac{\text{Cov}(u, \Theta)}{\text{Var}(\Theta)}. \quad (\text{A.17})$$

Thus, in equilibrium (expectations coincide with realizations) we obtain that

$$\gamma_U = \frac{1 - D\delta^*}{\lambda_U + \rho \text{Var}(u|\Theta)}, \quad (\text{A.18})$$

and $\mu_U = 0$, such that $Y = -\gamma_U p + \mu_U$.

Appendix B Residual supply curves

In this Appendix, following Kyle (1989), we illustrate the calculation of p_U , λ_U , p_I , and λ_I . Using the market clearing condition, each trader can compute his/her residual supply curve. Informed traders rearrange equation (1), so that

$$(N - 1)X + X + MY + ZQ = 0. \quad (\text{B.1})$$

Using the formulation of the demand schedule, X , described previously, equation (B.1) becomes:

$$(N - 1)(-\gamma_I p + \beta\iota + \mu_I) + X + M(-\gamma_U p + \mu_U) + ZQ = 0. \quad (\text{B.2})$$

Now, equation (B.2) can be rearranged as follows,

$$p = \frac{1}{(N - 1)\gamma_I + M\gamma_U} X + \frac{(N - 1)(\beta\iota + \mu_I) + M\mu_U + ZQ}{(N - 1)\gamma_I + M\gamma_U}, \quad (\text{B.3})$$

with

$$\lambda_I = \frac{1}{(N - 1)\gamma_I + M\gamma_U} \quad (\text{B.4})$$

and

$$p_I = \frac{(N - 1)(\beta\iota + \mu_I) + M\mu_U + ZQ}{(N - 1)\gamma_I + M\gamma_U}, \quad (\text{B.5})$$

such that $p = p_I + \lambda_I X$ and $\lambda_I = \frac{\partial p}{\partial X}$. Uninformed traders rearrange equation (1), so that

$$NX + (M - 1)Y + Y + ZQ = 0. \quad (\text{B.6})$$

Thus, similar to informed traders, they obtain:

$$\lambda_U = \frac{1}{N\gamma_I + (M - 1)\gamma_U} \quad (\text{B.7})$$

and

$$p_U = \frac{N(\beta\iota + \mu_I) + (M - 1)\mu_U + ZQ}{N\gamma_I + (M - 1)\gamma_U}, \quad (\text{B.8})$$

such that $p = p_U + \lambda_U Y$ and $\lambda_U = \frac{\partial p}{\partial Y}$.

Appendix C Price efficiency

In this Appendix, we derive the formula for price efficiency. We start from the fact that p and Θ (see [Appendix A](#) for the extraction of Θ) are informationally equivalent and proceed as follows:

$$Var(u|p)^{-1} = Var(u|\Theta)^{-1} = \left[Var(u) - \frac{Cov^2(u, \Theta)}{Var(\Theta)} \right]^{-1}. \quad (C.1)$$

First, we have that,

$$Var(u) = \sigma_u^2. \quad (C.2)$$

Also, we can easily derive that

$$Cov^2(u, \Theta) = \sigma_u^4. \quad (C.3)$$

Lastly, using equation [\(A.12\)](#), we can calculate $Var(\Theta)$ as follows:

$$Var(\Theta) = Var\left(\iota + \frac{ZQ}{N\beta}\right) \quad (C.4)$$

$$= Var(\iota) + Var\left(\frac{ZQ}{N\beta}\right) \quad (C.5)$$

$$= \sigma_u^2 + \sigma_e^2 + \frac{Z^2\sigma_Q^2}{N^2\beta^2} \quad (C.6)$$

$$= \sigma_u^2 + \sigma_e^2 + \frac{Z^2\sigma_Q^2}{\left(\frac{N}{\frac{\lambda_I}{\delta} + \rho\sigma_e^2}\right)^2}. \quad (C.7)$$

Adding all the pieces together, we have that

$$Var(u|p) = \sigma_u^2 - \frac{\sigma_u^4}{\sigma_u^2 + \sigma_e^2 + \left(\frac{Z}{\frac{N}{\frac{\lambda_I}{\delta} + \rho\sigma_e^2}}\right)^2 \sigma_Q^2}. \quad (C.8)$$

After rearranging the above equation, we obtain the final formula for price efficiency, represented by equation [\(3\)](#).

Appendix D Proofs of propositions (3.1) and (3.2)

Proof. Proposition (3.1)

Case 1: Assume that $N_{HFT} + Z_{HFT} > 0$ and $N_{non-HFT} + Z_{non-HFT} > 0$. We prove (11) and observe that (12) is actually an exact replica. We have,

$$\begin{aligned} \text{Var}^{-1}(u|p) &\geq \text{Var}^{-1}(u|p_{HFT}) \Leftrightarrow \\ \text{Var}(u|p) &\leq \text{Var}(u|p_{HFT}) \Leftrightarrow \\ \sigma_u^2 - \frac{\sigma_u^4}{\sigma_u^2 + \sigma_e^2 + \left(\frac{Z}{N^2\beta^2}\right)^2 \sigma_Q^2} &\leq \sigma_u^2 - \frac{N_{HFT}^2\beta^2\sigma_u^4}{N_{HFT}^2\beta^2(\sigma_u^2 + \sigma_e^2) + Z_{HFT}^2\sigma_Q^2} \Leftrightarrow \\ Z_{HFT} &\geq \frac{N_{HFT}Z}{N}, \end{aligned}$$

and the proof is completed. Inequality (13) follows along the same steps. Indeed, we have

$$\begin{aligned} \text{Var}^{-1}(u|p_{HFT}) &\geq \text{Var}^{-1}(u|p_{non-HFT}) \Leftrightarrow \\ \text{Var}(u|p_{HFT}) &\leq \text{Var}(u|p_{non-HFT}) \Leftrightarrow \\ \sigma_u^2 - \frac{N_{HFT}^2\beta^2\sigma_u^4}{N_{HFT}^2\beta^2(\sigma_u^2 + \sigma_e^2) + Z_{HFT}^2\sigma_Q^2} &\leq \sigma_u^2 - \frac{N_{non-HFT}^2\beta^2\sigma_u^4}{N_{non-HFT}^2\beta^2(\sigma_u^2 + \sigma_e^2) + Z_{non-HFT}^2\sigma_Q^2}. \end{aligned}$$

Simplifying and rearranging terms finally yields the desired result,

$$Z_{non-HFT}N_{HFT} \geq Z_{HFT}N_{non-HFT}.$$

Case 2: Assume that $N_{HFT} = Z_{HFT} = 0$, $M_{HFT} > 0$, and $N_{non-HFT} + Z_{non-HFT} > 0$. In this case all HFTs are uninformed and thus

$$\text{Var}(u|p_{HFT})^{-1} = \text{Var}^{-1}(u|p). \quad (\text{D.1})$$

The fact that $\text{Var}(u|p_{non-HFT})^{-1} = \text{Var}^{-1}(u|p)$ follows directly from the fact that $N_{non-HFT} = N$ and $Z_{non-HFT} = Z$.

Case 3: Assume that $N_{HFT} + Z_{HFT} > 0$, $N_{non-HFT} = Z_{non-HFT} = 0$ and $M_{non-HFT} > 0$. This case is symmetrical to case 2, and therefore the proof is completed. \square

Proof. Proposition (3.2)

For the first case, assume that $N_{HFT} = 0$ so that $N_{non-HFT} = N$ and $Z_{HFT} > 0$ (from case 1 of proposition (3.1)). Then, in view of (13), we get $0 > N_{non-HFT}Z_{HFT}$ which is a contradiction. Therefore, $N_{HFT} > 0$.

For the second case, in view of the assumption that $N_{HFT} + Z_{HFT} > 0$ and (11), in case 1 of proposition (3.1), we get that if $N_{HFT} = 0$, then $Z_{HFT} < 0$, which is a contradiction, and hence $N_{HFT} > 0$. □

Appendix E Summary statistics

Table E1: Summary statistics for $R_{cc}^{FULLBOOK}$ and $R_{ct}^{FULLBOOK}$ returns.

Return	Mean	Median	St.D.	Min.	Max.	Skewness	Kurtosis
$R_{close-to-close}$	0.0007	0.0007	0.0159	-0.1256	0.1242	0.0814	6.5331
$R_{close-to-08:31}$	-0.0010	0.0000	0.0092	-0.0778	0.0535	-0.5090	8.2849
$R_{close-to-08:32}$	-0.0010	-0.0001	0.0091	-0.0662	0.0526	-0.4087	7.4534
$R_{close-to-08:33}$	-0.0010	-0.0001	0.0091	-0.0680	0.0526	-0.3766	7.3212
$R_{close-to-08:34}$	-0.0010	-0.0001	0.0091	-0.0772	0.0522	-0.3488	7.3456
$R_{close-to-08:35}$	-0.0011	-0.0001	0.0092	-0.1065	0.0614	-0.4861	9.1931
$R_{close-to-08:36}$	-0.0011	-0.0001	0.0092	-0.1065	0.0594	-0.5050	9.0092
$R_{close-to-08:37}$	-0.0011	-0.0001	0.0093	-0.1065	0.0503	-0.5367	8.8378
$R_{close-to-08:38}$	-0.0011	-0.0001	0.0093	-0.1267	0.0515	-0.6211	10.4472
$R_{close-to-08:39}$	-0.0011	-0.0001	0.0094	-0.1267	0.0515	-0.6253	10.3930
$R_{close-to-08:40}$	-0.0011	-0.0001	0.0094	-0.1267	0.0492	-0.6563	10.3323
$R_{close-to-08:41}$	-0.0011	-0.0001	0.0095	-0.1267	0.0478	-0.7473	10.5286
$R_{close-to-08:42}$	-0.0011	0.0000	0.0095	-0.1267	0.0480	-0.7567	10.1624
$R_{close-to-08:43}$	-0.0010	0.0000	0.0095	-0.1267	0.0480	-0.7522	10.3300
$R_{close-to-08:44}$	-0.0010	0.0000	0.0096	-0.1267	0.0515	-0.7114	10.1775
$R_{close-to-08:45}$	-0.0010	0.0000	0.0096	-0.1267	0.0495	-0.7099	10.2967
$R_{close-to-08:46}$	-0.0009	0.0000	0.0093	-0.1267	0.0495	-0.6528	10.3401
$R_{close-to-08:47}$	-0.0009	0.0000	0.0092	-0.1267	0.0495	-0.6133	9.9155
$R_{close-to-08:48}$	-0.0008	0.0000	0.0092	-0.1267	0.0499	-0.6063	10.0114
$R_{close-to-08:49}$	-0.0008	0.0000	0.0092	-0.1267	0.0495	-0.5861	9.8928
$R_{close-to-08:50}$	-0.0008	0.0000	0.0092	-0.0622	0.0495	-0.4183	6.5838
$R_{close-to-08:51}$	-0.0004	0.0000	0.0110	-0.0735	0.0708	-0.2654	5.9887
$R_{close-to-08:52}$	-0.0004	0.0000	0.0109	-0.0611	0.0582	-0.2596	5.7431
$R_{close-to-08:53}$	-0.0003	0.0000	0.0108	-0.0727	0.0580	-0.2509	5.6558
$R_{close-to-08:54}$	-0.0002	0.0000	0.0107	-0.0727	0.0571	-0.2308	5.5955
$R_{close-to-08:55}$	-0.0003	0.0000	0.0105	-0.0727	0.0626	-0.2402	5.6494
$R_{close-to-08:56}$	-0.0003	0.0000	0.0105	-0.0978	0.0673	-0.4039	6.4916
$R_{close-to-08:57}$	0.0000	0.0000	0.0103	-0.1074	0.0626	-0.4596	7.1154
$R_{close-to-08:58}$	0.0000	0.0000	0.0098	-0.1029	0.0652	-0.4714	7.9761
$R_{close-to-08:59}$	0.0003	0.0001	0.0095	-0.0961	0.0678	-0.5554	9.5221
$R_{close-to-open}$	0.0004	0.0004	0.0080	-0.0938	0.0935	-0.6877	19.8954

Note: This Table reports summary statistics about the distribution of the logarithmic returns used in the regression analysis in equation (16), for the FULL BOOK case. For each time interval, we pool the returns across days and stocks. From the left column to the right column, we report the following: mean, median, standard deviation, minimum, maximum, skewness, and kurtosis.

Table E2: Summary statistics for R_{ct}^{HFT} returns.

Return	Mean	Median	St.D.	Min.	Max.	Skewness	Kurtosis
$R_{close-to-08:31}$	0.0011	0.0006	0.0166	-0.2548	0.1500	-2.2432	36.3416
$R_{close-to-08:32}$	0.0011	0.0006	0.0165	-0.2548	0.1500	-2.2402	37.0161
$R_{close-to-08:33}$	0.0011	0.0006	0.0162	-0.2548	0.1500	-2.3773	39.5372
$R_{close-to-08:34}$	0.0011	0.0006	0.0158	-0.2548	0.1500	-2.1634	37.8078
$R_{close-to-08:35}$	0.0010	0.0006	0.0156	-0.2548	0.1500	-2.2772	40.0822
$R_{close-to-08:36}$	0.0011	0.0006	0.0153	-0.2548	0.1500	-2.6458	46.3244
$R_{close-to-08:37}$	0.0011	0.0007	0.0152	-0.2548	0.1143	-2.7148	46.2233
$R_{close-to-08:38}$	0.0011	0.0006	0.0151	-0.2548	0.1143	-2.5329	44.4735
$R_{close-to-08:39}$	0.0011	0.0006	0.0150	-0.2548	0.1143	-2.5470	45.1119
$R_{close-to-08:40}$	0.0011	0.0006	0.0146	-0.2574	0.1143	-2.0815	39.8001
$R_{close-to-08:41}$	0.0012	0.0007	0.0140	-0.2106	0.1143	-1.2789	27.7186
$R_{close-to-08:42}$	0.0014	0.0007	0.0136	-0.2106	0.1143	-0.9581	26.7434
$R_{close-to-08:43}$	0.0014	0.0007	0.0134	-0.2106	0.1143	-0.9864	27.5997
$R_{close-to-08:44}$	0.0015	0.0007	0.0136	-0.2106	0.1143	-0.7953	26.1124
$R_{close-to-08:45}$	0.0014	0.0007	0.0136	-0.2106	0.1143	-0.8299	26.2023
$R_{close-to-08:46}$	0.0014	0.0006	0.0125	-0.2106	0.1143	-1.1641	32.9557
$R_{close-to-08:47}$	0.0014	0.0006	0.0125	-0.2106	0.1143	-1.2206	33.1967
$R_{close-to-08:48}$	0.0013	0.0006	0.0122	-0.2106	0.1143	-0.5776	25.1184
$R_{close-to-08:49}$	0.0012	0.0005	0.0122	-0.2106	0.1143	-0.5695	25.2796
$R_{close-to-08:50}$	0.0012	0.0005	0.0121	-0.2106	0.1143	-0.7190	25.4369
$R_{close-to-08:51}$	0.0026	0.0015	0.0189	-0.1030	0.1430	0.1276	5.8210
$R_{close-to-08:52}$	0.0027	0.0017	0.0189	-0.1219	0.1430	0.1239	5.8650
$R_{close-to-08:53}$	0.0027	0.0017	0.0190	-0.1772	0.1430	0.0170	6.7588
$R_{close-to-08:54}$	0.0027	0.0017	0.0190	-0.1772	0.1430	-0.0405	7.2928
$R_{close-to-08:55}$	0.0022	0.0013	0.0188	-0.1631	0.1430	-0.0304	7.6211
$R_{close-to-08:56}$	0.0021	0.0011	0.0183	-0.1631	0.1205	-0.0262	7.3154
$R_{close-to-08:57}$	0.0022	0.0014	0.0171	-0.1640	0.1464	-0.0378	9.3778
$R_{close-to-08:58}$	0.0014	0.0011	0.0131	-0.1112	0.1464	0.0628	10.1862
$R_{close-to-08:59}$	0.0012	0.0011	0.0116	-0.1043	0.1464	-0.1116	12.3201
$R_{close-to-open}$	0.0009	0.0008	0.0093	-0.1604	0.1770	0.4294	49.2367

Note: This Table reports summary statistics about the distribution of the logarithmic returns used in the regression analysis in equation (16), for the HFT case. For each time interval, we pool the returns across days and stocks. From the left column to the right column, we report the following: mean, median, standard deviation, minimum, maximum, skewness, and kurtosis.

Table E3: Summary statistics for $R_{ct}^{non-HFT}$ returns.

Return	Mean	Median	St.D.	Min.	Max.	Skewness	Kurtosis
$R_{close-to-08:31}$	-0.0013	-0.0001	0.0090	-0.0809	0.0549	-0.7397	9.1071
$R_{close-to-08:32}$	-0.0013	-0.0001	0.0090	-0.0735	0.0549	-0.6944	8.4655
$R_{close-to-08:33}$	-0.0014	-0.0002	0.0090	-0.0735	0.0549	-0.7060	8.3833
$R_{close-to-08:34}$	-0.0015	-0.0003	0.0091	-0.0757	0.0549	-0.7203	8.3459
$R_{close-to-08:35}$	-0.0016	-0.0003	0.0092	-0.0757	0.0540	-0.7244	8.2902
$R_{close-to-08:36}$	-0.0016	-0.0003	0.0092	-0.0757	0.0545	-0.7262	8.1660
$R_{close-to-08:37}$	-0.0016	-0.0003	0.0092	-0.0757	0.0538	-0.7423	8.1302
$R_{close-to-08:38}$	-0.0016	-0.0003	0.0092	-0.0757	0.0540	-0.7109	8.0980
$R_{close-to-08:39}$	-0.0016	-0.0003	0.0093	-0.0757	0.0564	-0.6997	8.0515
$R_{close-to-08:40}$	-0.0016	-0.0003	0.0093	-0.0757	0.0573	-0.7053	7.9797
$R_{close-to-08:41}$	-0.0016	-0.0003	0.0093	-0.0757	0.0571	-0.7618	8.1018
$R_{close-to-08:42}$	-0.0016	-0.0003	0.0094	-0.0757	0.0571	-0.7605	7.9985
$R_{close-to-08:43}$	-0.0016	-0.0003	0.0094	-0.0757	0.0573	-0.7653	7.9925
$R_{close-to-08:44}$	-0.0016	-0.0003	0.0095	-0.0757	0.0573	-0.7821	8.1974
$R_{close-to-08:45}$	-0.0016	-0.0003	0.0095	-0.0757	0.0573	-0.7720	8.0936
$R_{close-to-08:46}$	-0.0016	-0.0003	0.0095	-0.0757	0.0573	-0.7466	7.9608
$R_{close-to-08:47}$	-0.0016	-0.0003	0.0095	-0.0757	0.0573	-0.7428	8.0014
$R_{close-to-08:48}$	-0.0016	-0.0003	0.0095	-0.0757	0.0587	-0.7299	7.9115
$R_{close-to-08:49}$	-0.0015	-0.0003	0.0095	-0.0757	0.0499	-0.7219	7.8485
$R_{close-to-08:50}$	-0.0015	-0.0003	0.0096	-0.0757	0.0517	-0.7857	8.0676
$R_{close-to-08:51}$	-0.0015	-0.0002	0.0098	-0.0773	0.0872	-0.7238	8.6603
$R_{close-to-08:52}$	-0.0015	-0.0003	0.0097	-0.0773	0.0614	-0.7940	8.1989
$R_{close-to-08:53}$	-0.0015	-0.0003	0.0098	-0.1103	0.0614	-0.9411	9.8114
$R_{close-to-08:54}$	-0.0015	-0.0003	0.0098	-0.1019	0.0614	-0.8764	9.4854
$R_{close-to-08:55}$	-0.0015	-0.0003	0.0098	-0.1100	0.0614	-0.9172	10.3991
$R_{close-to-08:56}$	-0.0015	-0.0002	0.0105	-0.1203	0.0614	-0.9541	10.4244
$R_{close-to-08:57}$	-0.0015	-0.0002	0.0107	-0.1203	0.0598	-0.7396	8.5024
$R_{close-to-08:58}$	-0.0014	-0.0002	0.0110	-0.1203	0.0742	-0.7415	9.0965
$R_{close-to-08:59}$	-0.0012	-0.0001	0.0113	-0.1272	0.0870	-0.6979	10.0538
$R_{close-to-open}$	-0.0010	0.0000	0.0110	-0.1203	0.0870	-0.6088	10.6753

Note: This Table reports summary statistics about the distribution of the logarithmic returns used in the regression analysis in equation (16), for the non-HFT case. For each time interval, we pool the returns across days and stocks. From the left column to the right column, we report the following: mean, median, standard deviation, minimum, maximum, skewness, and kurtosis.